# Dynamic reliability analysis using the extended support vector regression (X-SVR)

Jinwen Feng[a]; Lei Liu[a]; Di Wu[a,*]; Guoyin Li[b]; Michael Beer[c,d,e]; Wei Gao[a,**]

[a]Centre for Infrastructure Engineering and Safety (CIES),
School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia

[b]School of Mathematics and Statistics, The University of New South Wales,
Sydney, NSW 2052, Australia

[c]Institute for Risk and Uncertainty, University of Liverpool, Liverpool L69 3GH, UK

[d]Institute for Risk and Reliability, Leibniz University Hannover, Germany

[e]International Joint Research Center for Engineering Reliability and Stochastic Mechanics (ERSM), Tongji University, China

## Abstract

For engineering applications, the dynamic system responses can be significantly affected by uncertainties in the system parameters including material and geometric properties as well as by uncertainties in the excitations. The reliability of dynamic systems is widely evaluated based on the first-passage theory. To improve the computational efficiency, surrogate models are widely used to approximate the relationship between the system inputs and outputs. In this paper, a new machine learning based metamodel, namely the *extended support vector regression* (X-SVR), is proposed for the reliability analysis of dynamic systems via utilizing the first-passage theory. Furthermore, the capability of X-SVR is enhanced by a new kernel function developed from the vectorized Gegenbauer polynomial, especially for solving complex engineering problems. Through the proposed approach, the relationship between the extremum of the dynamic responses and the input uncertain parameters is approximated by training the X-SVR model such that the probability of failure can be efficiently predicted without using other computational tools for numerical analysis, such as the finite element analysis (FEM). The feasibility and performance of the proposed surrogate model in dynamic reliability analysis is investigated by comparing it with the conventional ε-insensitive support vector regression (ε-SVR) with Gaussian kernel and Monte Carlo simulation (MSC). Four numerical examples are adopted to evidently demonstrate the practicability and efficiency of the proposed X-SVR method.

## Keywords:

Extended support vector regression; Generalized Gegenbauer kernel; Dynamic analysis; Reliability analysis; Surrogate model

* Corresponding author. Email: di.wu@unsw.edu.au
** Corresponding author. Email: w.gao@unsw.edu.au

## 1. Introduction

The uncertain parameters, such as material properties, dimensions of the engineering product, and loading regimes, are inherently associated with the practical engineering systems and may lead to considerable fluctuations in the dynamic responses. The dynamic reliability of engineering systems is essential and important to be investigated so that the effects of the uncertain variables can be thoroughly evaluated in the analyses and designs [1-4]. Generally, stochastic approaches are applied in the reliability analysis by modeling the uncertainties as random variables/fields with statistical information (i.e. mean and standard deviation) [5-9]. As a matter of fact, the corresponding stochastic response in dynamics is time-dependent and should therefore be represented by a stochastic process, which increases the computational cost in comparison with the static reliability analysis.

In the past decades, there are numerous methods that have been developed for estimating the dynamic reliability of engineering systems, which focus on estimating the first-passage probability by evaluating the mean out-crossing rate [10]. The integration of the out-crossing rate is usually based on considering the out-crossing events either individually (Poisson model) or in a group (Markov model) [11]. In addition to the stochastic approaches, a non-probabilistic alternative convex process model is introduced in [12] to solve the first-passage reliability analysis. The most efficient stochastic approaches in this regard are based on approximately determining the probability density function (PDF) of the extreme system performance, which allows for a direct evaluation and estimation of the failure probability. The extreme system performance is quantified with an extreme value distribution (EVD) for the first-passage reliability computation [11, 13]. In the EVD approach, the time-dependent reliability analysis is beneficially converted into the time-independent reliability evaluation. For the problems with small variations of coefficients, an envelope function method was introduced with the first-order approximation of the motion error function [14]. Concepts of stochastic averaging/linearization [15], of dimension reduction [16], and of numerical path integral solutions [17] provide currently the most efficient pathways to solve the first passage problem. Nevertheless, the derivation of a closed-form equation for the extreme values is technically difficult for generalized dynamic responses [18]. Despite of the comprehensively established theory, the first-passage probability of failure can be analytically obtained only in limited cases and is mostly restricted to single degree of freedom (SDOF) problems [19]. The consideration of nonlinearities and the expansion to several degrees of freedom are topics of current research [20].

As an alternative pathway, the probability distribution of extreme system performance can be obtained via sampling-based approaches. Within the framework of EVD, the probability density evolution method [21] and the equivalent extreme value approach [22] are developed for estimating

the probability density function of the extreme values in the responses of the dynamic systems numerically. However, the numerical integration for evaluation of the PDF using these approaches still requires a large number of deterministic dynamic analyses with respect to selected representative points of input random variables. To increase the computational efficiency, the numerical integration can be replaced by using the first-order reliability method (FORM) in association with some specific adjustments such as PHI2 method [23] or discretized stochastic processes [11]. Although useful for small practical cases, FORM is associated with its known limitations (only weak nonlinearities, small dimensionality etc.). Also, the advanced sampling schemes have been proven powerful in enhancing the efficiency of the generally applicable Monte Carlo simulation techniques [24].

Surrogate models have been widely employed as computationally highly efficient approach for reliability assessments and uncertainty quantifications. The relationship between the input variables and structural outputs is approximated by the metamodeling approach, which is an explicit formulation. This leads to the efficient evaluation of the limit state function. Via utilizing the Kriging technique, or so-called Gaussian process model, a nested extreme response surface (NERS) method was developed to efficiently identify the extreme time responses as the dynamic reliability analysis can be conducted using static reliability analysis methods [25]. Other Kriging-based reliability analysis approaches for dynamics are also developed for various engineering applications [26, 27]. Recently, a Chebyshev method was proposed for the dynamic uncertainty analysis of multibody mechanical systems and extended for analyzing the dynamic responses of structures with uncertain variables [28, 29]. In the past decades, machine learning techniques, such as the support vector machine (SVM), have been extensively studied in the structural reliability assessments. For example, the least squares support vector machine has been utilized for the dynamic reliability analysis of turbomachinery in [30]. Unlike the surrogate models based on the polynomial chaos expansion (PCE), the SVM is capable to bypass the curse of dimensionality and can also handle nonlinear problems effectively.

In this paper, a novel surrogate model, namely the *extended support vector regression* (X-SVR) is proposed for the dynamic reliability analysis. The underpinned analysis framework is based on the first-passage failure theorem. Comparing with the classical SVM approach, the satisfaction of the Mercer's condition is not prerequisite for the kernel functions employed in the nonlinear X-SVR. To further enhance the capability of the kernelized X-SVR in approximating complex functions, a new orthogonal polynomial kernel function, namely the *generalized Gegenbauer kernel* (GGK), is proposed. The introduced GGK is an admissible Mercer kernel, and it can be applied to other kernel learning methods which strictly require the satisfaction of the Mercer's condition. As a mixed kernel function, the proposed GGK consists of both orthogonal polynomial and Gaussian kernel function.

Therefore, the advantages of both global and local kernels are included. Instead of using the conventional grid search technique, the hyperparameters of the X-SVR model are efficiently selected by the Bayesian optimization algorithm. After the establishment of the X-SVR surrogate model with GGK, the limit state function based on the first-passage principle can be explicitly approximated from the obtained X-SVR regression function. Subsequently, the probability of failure can be evaluated by Monte-Carlo Simulation (MCS) method with the constructed metamodel instead of using the actual computationally expensive numerical models (e.g., finite element analysis) with high-fidelity. This leads to a tremendous reduction of the computational cost.

The rest of the paper is organized as follows. Section 2 presents a brief review on the theoretical background of the dynamic reliability analysis using the first-passage probability. Then, Section 3 demonstrates the detailed derivations of the proposed extended support vector machine and the *generalized Gegenbauer kernel* (GGK) function. Next, the procedures for applying the proposed X-SVR on dynamic reliability analysis are introduced in Section 4. The capability of the X-SVR with GGK is verified against two benchmark problems and two reliability analysis problems in Section 5. Finally, some concluding remarks are presented in Section 6.

## 2. The dynamic reliability analysis

### *2.1 Stochastic dynamic response of structure*

The global equations of motion for a linear engineering structure with multi-degrees of freedom (MDOF) can be expressed as:

$$\mathbf{M}_s\ddot{\mathbf{u}}(t)+\mathbf{C}_s\dot{\mathbf{u}}(t)+\mathbf{K}_s\mathbf{u}(t)=\mathbf{F}(t) \tag{1}$$

where $\mathbf{M}_s$, $\mathbf{C}_s$ and $\mathbf{K}_s$ are the mass, damping and stiffness matrices of the structure, respectively; $\mathbf{F}(t)$ denotes the external excitation vector which is time-dependent; $\ddot{\mathbf{u}}(t)$, $\dot{\mathbf{u}}(t)$ and $\mathbf{u}(t)$ are the time dependent acceleration, velocity and displacement vectors, respectively. Due to the existence of uncertain parameters, the mass, damping, stiffness matrices and the external excitation are non-deterministic [31-33]. In this study, the uncertainties are considered as independent random variables and $\mathbf{M}_s$, $\mathbf{C}_s$ and $\mathbf{K}_s$ can be expressed as functions of the random parameters. Without loss of generality, the random vector $\mathbf{x}\in\Re^n$ is adopted as the collection of uncertainties included in both parameters in the dynamic system and the external excitation. Thus, the non-deterministic dynamic responses can be calculated by the following global equation:

$$\mathbf{M}_s(\mathbf{x})\ddot{\mathbf{u}}(t)+\mathbf{C}_s(\mathbf{x})\dot{\mathbf{u}}(t)+\mathbf{K}_s(\mathbf{x})\mathbf{u}(t)=\mathbf{F}(\mathbf{x},t) \tag{2}$$

Given a time interval $[0,T]$, the initial condition is considered as deterministic as Eq. (3):

$$\mathbf{u}(0) = \mathbf{u}_0,\ \dot{\mathbf{u}}(0) = \dot{\mathbf{u}}_0 \tag{3}$$

The stochastic dynamic response for a non-defect system is a stochastic process which is dependent on the random vector $\mathbf{x}$. Thus, the solution of Eq. (2) can be conveniently expressed as:

$$\mathbf{u}(t) = \mathbf{H}(\mathbf{x}, t) \tag{4}$$

where $\mathbf{H}$ denotes a deterministic operator. Despite of the existence of Eq. (4), the explicit formulation of $\mathbf{H}(\mathbf{x}, t)$ is usually available for some special cases rather than practical engineering problems with MDOF [10, 21]. Subsequently, the determination of an explicit tractable expression for the joint probability density function (PDF) of $\mathbf{u}(t)$ becomes computationally infeasible. Among the numerical approaches developed for approximately evaluating the statistical characteristics of stochastic dynamic response of structures, the Monte Carlo method is considered as the versatile strategy and widely adopted due to the straightforward implementation process [34, 35].

*2.2 The first-passage failure theorem*

The reliability of a system is typically evaluated by calculating the probability of failure which is commonly measured by the responses, such as the stresses, strains or displacements of the specified critical element or control point in structural dynamics. The systems are considered as unsafe if the concerned responses exceed the safety threshold for the first time. Mathematically, the computation of the probability of failure ( $p_f$ ) is expressed as

$$p_f = \Pr\{g(\mathbf{x}) \le 0\} = \int_{g(\mathbf{x})\le 0} f_{\mathbf{x}}^{PDF}(\mathbf{x})d\mathbf{x} = \int_{\Re^n} \mathrm{I}[g(\mathbf{x}) \le 0] f_{\mathbf{x}}^{PDF}(\mathbf{x})d\mathbf{x} \tag{5}$$

where $\Pr\{\bullet\}$ denotes the probability; $\mathbf{x} \in \Re^n$ is the input random vector; $\mathrm{I}[\bullet]$ is the indicator function which equals to 1 if $[\bullet]$ is "true" and 0 when $[\bullet]$ is false"; $f_{\mathbf{x}}^{PDF}(\mathbf{x})$ denotes the joint PDF for $\mathbf{x}$ and $g(\mathbf{x})$ represents the limit state function, which defines a structural failure when $g(\mathbf{x}) \le 0$. Within the context of structural dynamic reliability, the structural outputs including displacement and stress become time-variant uncertainties. Thus, the limit state function can be explicitly expressed as a function of random variable $\mathbf{x}$ and time $t$. For a given time interval $[0, T]$, the probability of failure can be described as

$$p_f = \Pr\{\exists t \in [0, T], g(\mathbf{x}, t) \le 0\} \tag{6}$$

Thus, an efficient and accurate evaluation of $p_f$ is the key task in the structural dynamic reliability analysis.

Conventionally, for structural dynamic reliability, the probability of failure is widely computed by adopting the first-passage theory which is developed based on stochastic process [36]. In general,

the security margin for the first passage theory can be categorized into single-sided barrier, two-sided barrier and enveloping barrier [37, 38]. Among them, the single-sided barrier can be regarded as a special case of the two-sided barrier problems. In this work, the two-sided barrier circumstance is studied for demonstrating the capability of the proposed X-SVR meta-model. The safe domain for two-sided barrier problem in a given time interval $[0,T]$, as shown in Fig. 1, is defined as $b_2 < u(\mathbf{x},t) < b_1$ which is equivalent to $g(\mathbf{x},t) > 0$, where $u(\mathbf{x},t)$ is the stochastic structural dynamic response; $b_1$ and $b_2$ are the upper and lower threshold, respectively.

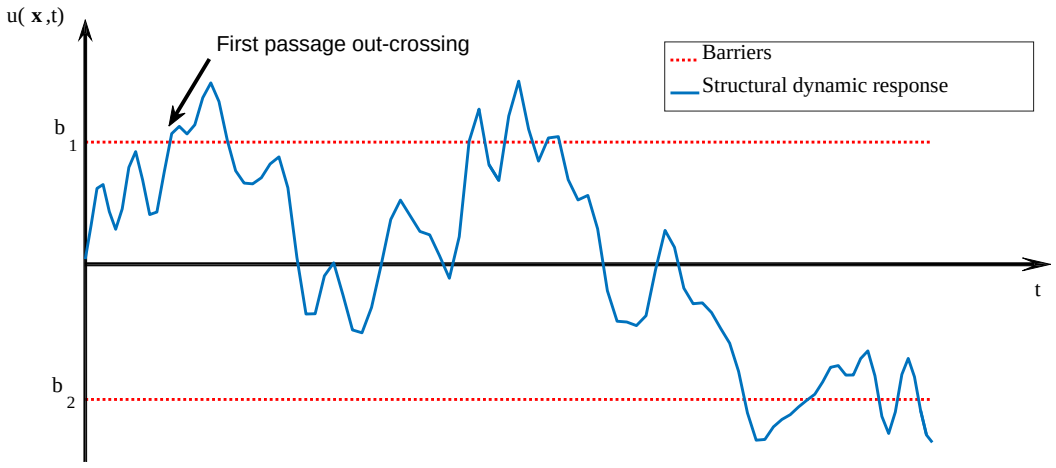



Fig. 1. The first-passage failure model for dynamic response of structure

For a structural system with input uncertain variable $\mathbf{X}$, the probability of failure based on the first-passage theory is obtained by evaluating the probability of the first occurrence of an excursion of the performance function (stress or displacement) exceeds the safe domain. In this context, probability of failure in time interval $[0,T]$ can be expressed as

$$p_f = \Pr\{\exists t \in [0,T], \max u(\mathbf{x},t) \geq b_1 \cup \min u(\mathbf{x},t) \leq b_2\} \tag{7}$$

For the two-sided barrier where $b_1 = -b_2 = b$ within the time interval $[0,T]$, the limit state function can be defined with the extreme of the dynamic response as $g(\mathbf{x}) = b - \max(|u(\mathbf{x},t)|)$. Thus, the $p_f$ expressed in Eq. (7) can be simplified as Eq. (8):

$$p_f = \Pr\{\exists t \in [0,T], b - \max(|u(\mathbf{x},t)|) \leq 0\} \tag{8}$$

By implementing the out-crossing rate-based approach, the first-passage failure probability is approximated according to the out-crossing rate $v^+(t)$ which is defined in Eq. (9):

$$v^+(t) = \lim_{\Delta t \to 0, \Delta t > 0} \frac{\Pr\{g(\mathbf{x},t) > 0 \cap g(\mathbf{x},t+\Delta t) < 0\}}{\Delta t} \tag{9}$$

Among the existing hypothetical models for computing the cumulative probability of failure based on the out-crossing rate, the Poisson and Markov models are the two widely adopted [11]. The

Poisson model assumes that the out-crossing events are mutually independent and the occurrence follows Poisson distribution. The Markov approach adopts an alternative path by assuming that the out-crossing events tend to occur in independent groups. Accordingly, the probability of failure in a given time interval $[0,T]$ evaluated by the Poisson and Markov models can be expressed in Eqs. (10) and (11), respectively:

$$p_f = 1 - \exp\left[ -\int_0^T v^+(t)dt \right] \tag{10}$$

$$p_f = 1 - \exp\left[ -\int_0^T \frac{v^+(t)}{1 - p_{f,ins}(t)} dt \right] \tag{11}$$

where $p_{f,ins}(t) = \Pr\{g(\mathbf{x},t) \le 0\}$ denotes the instantaneous probability of failure at time $t$. It can be observed from Eqs. (10) and (11) that both approaches require the integration of the out-crossing rate which is difficult to be obtained for the general stochastic process [19]. Thus, the closed-form solutions for Eqs. (6) - (8) are usually available for rather simple and special problems.

In this context, the Monte Carlo simulation (MCS) is commonly employed for computing the estimated probability of failure $\hat{p}_f$ by generating a large number of samples [12, 34]. Given $m$ samples $\mathbf{x}_i$ ($i = 1,2,...,m$) for input variables, the probability of failure can be approximated by Eq. (12):

$$p_f \approx \hat{p}_f = \frac{1}{m} \sum_{i=1}^{m} \mathrm{I}[g(\mathbf{x}_i) \le 0] = \frac{m_f}{m} \tag{12}$$

where $m_f$ is the number of samples that result in the failure of the structure. In Eq. (8), the limit state function can represent either internal force or structural deformation, which leads the first-passage failure mode to either the strength failure or deformation failure criterion [12]. Despite the fact that the first-passage theorem is conceptually simple through the MCS approach, the determination of the first-passage probability requires the computation of the whole dynamic response in the given time interval $[0,T]$ recursively. Thus, the majority of the computational cost is spent on repeatedly evaluating the limit state functions not to mention that the finite element analysis for complex structures in each simulative cycle can be computationally intensive [39, 40]. Alternatively, the meta-modelling techniques are introduced to approximate the relationship between inputs and outputs by an explicit function (i.e. polynomial). The meta-models are generally much less complicated than the original structural models, and it is expected that the computing effort will be reduced by approximating the limit state function by using surrogate models.

## 3. The extended support vector regression (X-SVR) with generalized Gegenbauer kernel (GGK)

In this section, a new surrogate model, namely the *extended support vector regression* (X-SVR), is proposed for reducing the computational cost of the conventional MCS approach. Furthermore, a new orthogonal polynomial kernel based on the Gegenbauer polynomial is introduced and adopted in the kernelized X-SVR. As a statistical learning method, the X-SVR model is developed as an extension of the doubly regularized support vector machine (DrSVM) which will be briefly described in Section 3.1. Then, the detailed formulation of the X-SVR and proposed orthogonal polynomial kernel will be presented.

### *3.1 The doubly regularized support vector machine (DrSVM)*

The aim of the SVM is to find a hyperplane which has the maximum distance to the closest data points located on each side. [41] Given the training dataset with input $\mathbf{x}_{train} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_i, ..., \mathbf{x}_m]^T \in \Re^{m \times n}$ ($\mathbf{x}_i \in \Re^n, i = 1, 2, ..., m$) and output $\mathbf{y}_{train} \in \Re^m$, the hyperplane that separating the two classes is then given as Eq. (13),

$$\hat{f}(\mathbf{x}) = \mathbf{w}^T \mathbf{x} - \gamma \tag{13}$$

where $m$ is the number of training samples; $n$ denotes the number of input variables; $\mathbf{x} \in \Re^n$ is the input random vector; $\mathbf{w} \in \Re^n$ denotes the normal to the hyperplane and $\gamma \in \Re$ denotes the bias. In the case of applying the Support Vector theory to regression estimation, namely Support Vector Regression (SVR) [42], the Eq. (13) is the regression function to be obtained and $y_{train}^i \in \Re$ is the output of the true function $f(\mathbf{x}_i)$. The linear SVR can be extended to nonlinear regression analysis by implicitly mapping the input data $\mathbf{x}_i$ from the low-dimension origin space $\Re^n$ into a higher-dimensional Euclidian space or even infinite dimensional Hilbert feature space $F$ by using an appropriate mapping function $\mathbf{\Phi}(\mathbf{x}_i)$. Thus, the feature space $F$ is also named as *intrinsic vector space* [43] and the mapping can be illustrated as:

$$\mathbf{x}_i = [x_{i,1}, x_{i,2}, ..., x_{i,n}]^T \mapsto \mathbf{\Phi}(\mathbf{x}_i) = [\phi^1(\mathbf{x}_i), \phi^2(\mathbf{x}_i), ..., \phi^{\tilde{\Gamma}}(\mathbf{x}_i)]^T \tag{14}$$

where the dimension $\tilde{\Gamma}$ is referred as the *intrinsic degree*, which can be either finite or infinite.

The coefficients $\mathbf{w}$ and $\gamma$ can be obtained by solving the following quadratic programming problem:

$$\min_{\mathbf{w},\gamma,\xi} : \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{m}(\xi_i + \hat{\xi}_i) \tag{15a}$$

$$s.t. \quad \begin{cases} y_{train}^{i} - (\mathbf{w}^T \boldsymbol{\Phi}(\mathbf{x}_i) - \gamma) \le \varepsilon + \xi_i \\ (\mathbf{w}^T \boldsymbol{\Phi}(\mathbf{x}_i) - \gamma) - y_{train}^{i} \le \varepsilon + \hat{\xi}_i \\ \xi_i, \hat{\xi}_i \ge 0, i = 1, 2, ..., m \end{cases} \tag{15b}$$

where $C > 0$ is the penalty constant which is defined for maintain a proper balance between the flatness of $\hat{f}(\mathbf{x})$ and the empirical error; $\|\bullet\|_2$ denotes the $l_2$ norm of $\bullet$; $\xi$ and $\hat{\xi}$ are the slack variables introduced for respectively allowing some excess positive and negative deviations for the prediction function $\hat{f}(x)$; $\varepsilon$ represents the tolerable deviation between $y_{train}^{i}$ and $\hat{f}(\mathbf{x}_i)$ which is defined by the so-called ε-insensitive loss function $l^{\varepsilon}(y_{train}^{i} - \hat{f}(\mathbf{x}_i))$ [42] as expressed in Eq. (16).

$$l^{\varepsilon}(y_{train}^{i} - \hat{f}(\mathbf{x}_i)) = \max(0, \left| y_{train}^{i} - \hat{f}(\mathbf{x}_i) \right| - \varepsilon), i = 1, ...m \tag{16}$$

Due to the employment of the ε-insensitive loss function, the regression method expressed by Eq. (15) is commonly referred as ε-SVR. Such error tolerance can be demonstrated in Fig. 2 by using a one-dimensional linear SVR model.

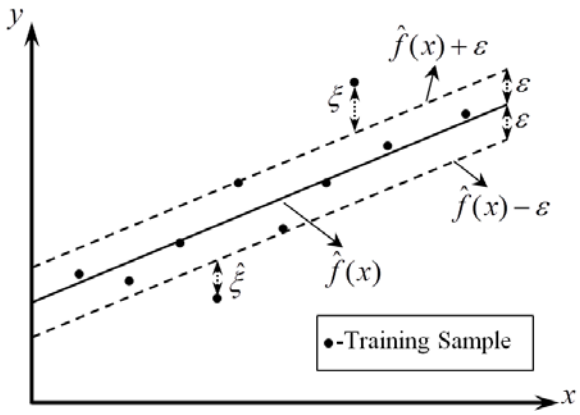



Fig. 2. The ε-insensitive band for a one-dimensional linear SVR

Then, the regression function in Eq. (13) is now expressed as Eq. (17):

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) K(\mathbf{x}_i, \mathbf{x}) - \gamma \tag{17}$$

where the $\alpha_i$ and $\hat{\alpha}_i$ are the Lagrange multipliers used for solving the optimization problem expressed in Eq. (15). The kernel functions used for support vector machine/ regression should satisfy the Mercer's theorem which requires $K(\mathbf{x}_i, \mathbf{x}_j)$ to be positive semi-definite [43]. This property is also a guarantee that the optimization problem expressed by Eq. (15) can be solved as a convex quadratic programming problem.

As an extension of the theory of support vector machine (SVM), the doubly regularized support vector machine (DrSVM) was proposed by [44] such that the classification and feature selection can

be conducted simultaneously. Theoretically, the DrSVM is a combination of elastic net penalty which contains both $l_1$ norm and $l_2$ norm penalty with the hinge loss function for reducing the effect of noise and outliers in the training dataset [45]. Accordingly, the DrSVM can be expressed as follows:

$$\min_{\mathbf{w},\gamma} : \frac{\lambda_1}{2}\|\mathbf{w}\|_2^2 + \lambda_2 \|\mathbf{w}\|_1 + \sum_{i=1}^{m}[1 - y_{train}^{i} f(\mathbf{x}_i)]_+ \tag{18}$$

where the tuning parameters $\lambda_1, \lambda_2 > 0$ control of balance between the classification performance and feature selection; $\|\bullet\|_1$ denotes the $l_1$ norm of $\bullet$; $[1 - y_{train}^{i} f(\mathbf{x}_i)]_+$ denotes the hinge loss function.

Due to the additional capability in feature selection, the DrSVM is attracting increasing attention since it was firstly emerged. Recently, a new DrSVM, namely the *pq*-SVM was proposed by [46] as an alternative approach for solving the optimization problem defined in Eq. (18). In the *pq*-SVM model, two non-negative variables $\mathbf{p}, \mathbf{q} \in \Re^n$ are introduced such that $\mathbf{w} = \mathbf{p} - \mathbf{q}$. The variables $\mathbf{p}$ and $\mathbf{q}$ are defined as in Eq. (19):

$$p_i := \begin{cases} 0, & w_i \le 0 \\ w_i, & w_i > 0 \end{cases} \text{ and } q_i := \begin{cases} -w_i, & w_i < 0 \\ 0, & w_i \ge 0 \end{cases} \tag{19}$$

It is indicated by the definition in Eq. (19) that $p_i q_i = 0$ is promised $\forall i$. Thus, the $\|\mathbf{w}\|_1$ and $\|\mathbf{w}\|_2^2$ can be alternatively expressed as Eqs (20) and (21):

$$\begin{aligned} \|\mathbf{w}\|_1 &= |w_1| + |w_2| + \ldots + |w_n| \\ &= p_1 + q_1 + p_2 + q_2 + \ldots + p_n + q_n \\ &= \mathbf{e}_n^T(\mathbf{p} + \mathbf{q}) \end{aligned} \tag{20}$$

$$\begin{aligned} \|\mathbf{w}\|_2^2 &= \|\mathbf{p} - \mathbf{q}\|_2^2 \\ &= \|\mathbf{p}\|_2^2 + \|\mathbf{q}\|_2^2 - 2\mathbf{p}^T\mathbf{q} \\ &= \|\mathbf{p}\|_2^2 + \|\mathbf{q}\|_2^2 \end{aligned} \tag{21}$$

where $\mathbf{e}_n = [1,1,\ldots,1]^T \in \Re^n$. Such decomposition is proved as an effective approach by the implementation in nonlinear optimization [47] and $l_1$ norm support vector regression [42]. In this context, the decision function (Eq. (13)) can be reformulated as:

$$f(\mathbf{x}) = \mathbf{x}^T(\mathbf{p} - \mathbf{q}) - \gamma \tag{22}$$

Similar as the soft margin SVM, a non-negative slack variable $\xi \in \Re^m$ is introduced into the optimization to control the marginal error. Therefore, the *pq*-SVM can be expressed as:

$$\min_{\mathbf{p},\mathbf{q},\gamma,\xi} : \frac{\lambda_1}{2}(\|\mathbf{p}\|_2^2 + \|\mathbf{q}\|_2^2) + \lambda_2 \mathbf{e}_n^T(\mathbf{p}+\mathbf{q}) + \mathbf{e}_m^T\boldsymbol{\xi} \tag{23a}$$

$$s.t. \quad \begin{cases} \mathbf{D}(\mathbf{x}_{train}(\mathbf{p}-\mathbf{q}) - \gamma \mathbf{e}_m^T) + \boldsymbol{\xi} \geq \mathbf{e}_m^T \\ \mathbf{p},\mathbf{q} \geq \mathbf{0}_n; \boldsymbol{\xi} \geq \mathbf{0}_m \end{cases} \tag{23b}$$

where $\mathbf{e}_m = [1,1,...,1]^T \in \Re^m$ and $\mathbf{D} \in \Re^{m \times m}$, as expressed in Eq. (24), is a diagonal matrix which contains all labels associated with training dataset

$$\mathbf{D} = \begin{bmatrix} y_{train}^1 & & \\ & \ddots & \\ & & y_{train}^m \end{bmatrix} \tag{24}$$

By using the Lagrange method, the *pq*-SVM can be reformulated into a quadratic programming problem [46], which indicates the advantage of the introduction of variables $\mathbf{p}$ and $\mathbf{q}$ for decomposing the $l_1$ norm of coefficient $\mathbf{w}$. In addition to the conventional classification problem, the *pq*-SVM is modified as knowledge-based SVM by incorporating the prior knowledge in the form of uncertain linear constraints [48]. Despite of the successful implementation of DrSVM and *pq*-SVM in classification problem, according to the authors' best knowledge, a doubly regularized support vector regression model has not yet been developed.

*3.2 The proposed extended support vector regression (X-SVR)*

Inspired by the success of *pq*-SVM in classification, a new support vector regression (SVR) model, namely the *extended support vector regression* (X-SVR) is developed in this study by adopting the concept of DrSVM and extended from binary classification to the regression estimation. In the proposed regression model, the decomposition method applied in the *pq*-SVM is adopted such that the $l_1$ norm computation $\|\mathbf{w}\|_1$ is eliminated. Additionally, instead of using the widely adopted linear ε-insensitive loss function expressed in Eq. (16), the proposed X-SVR incorporated the quadratic ε-insensitive loss function which is defined in Eq. (25)

$$l_2^{\varepsilon}(y_{train}^i - \hat{f}(\mathbf{x}_i)) = \left| y_{train}^i - \hat{f}(\mathbf{x}_i) \right|^2, i = 1,...m \tag{25}$$

By using the non-negative slack variables $\boldsymbol{\xi}, \hat{\boldsymbol{\xi}} \in \Re^m$, the X-SVR can be derived by solving the constrained optimization problem formulated as Eq. (26):

$$\min_{\mathbf{p},\mathbf{q},\gamma,\xi,\hat{\xi}} : \frac{\lambda_1}{2}(\|\mathbf{p}\|_2^2 + \|\mathbf{q}\|_2^2) + \lambda_2 \mathbf{e}_n^T(\mathbf{p}+\mathbf{q}) + \frac{C}{2}(\boldsymbol{\xi}^T\boldsymbol{\xi} + \hat{\boldsymbol{\xi}}^T\hat{\boldsymbol{\xi}}) \tag{26a}$$

$$s.t. \quad \begin{cases} \mathbf{x}_{train}(\mathbf{p}-\mathbf{q}) - \gamma \mathbf{e}_m^T - \mathbf{y}_{train} \le \varepsilon \mathbf{e}_m^T + \boldsymbol{\xi} \\ \mathbf{y}_{train} - (\mathbf{x}_{train}(\mathbf{p}-\mathbf{q}) - \gamma \mathbf{e}_m^T) \le \varepsilon \mathbf{e}_m^T + \hat{\boldsymbol{\xi}} \\ \mathbf{p}, \mathbf{q} \ge \mathbf{0}_n; \boldsymbol{\xi}, \hat{\boldsymbol{\xi}} \ge \mathbf{0}_m \end{cases} \tag{26b}$$

where $C > 0$ is the penalty constant (so-called box constraint) which controls the emphasis on the error minimization. As pointed out in [49], the introduction of quadratic ε-insensitive loss function leads to the redundancy of the non-negative constraint for the slack variables $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}$. Furthermore, $\gamma^2$ (the square of bias parameter) is added to the objective function, which enables the simultaneous optimizing the orientation and location of the regression model [46, 50]. For the sake of simplicity, Eq. (26) can be further modified as the following optimization problem:

$$\min_{\mathbf{z},\gamma} : \frac{1}{2}(\mathbf{z}^T \hat{\mathbf{C}} \mathbf{z} + \gamma^2) + \lambda_2 \mathbf{b}^T \mathbf{z} \tag{27a}$$

$$s.t. \quad (\hat{\mathbf{A}} + \mathbf{I}_{(2m+2n)\times(2m+2n)})\mathbf{z} + (\varepsilon \mathbf{I}_{(2m+2n)\times(2m+2n)} + \gamma \hat{\mathbf{G}})\hat{\mathbf{e}} + \hat{\mathbf{d}} \ge \mathbf{0}_{2m+2n} \tag{27b}$$

where $\mathbf{I}_{(2m+2n)\times(2m+2n)} \in \Re^{(2m+2n)\times(2m+2n)}$ denotes the identity matrix and $\mathbf{0}_{2m+2n} \in \Re^{2m+2n}$ denotes the zero vector. The matrices $\hat{\mathbf{C}}$, $\hat{\mathbf{G}}$ and $\hat{\mathbf{A}}$ are defined as:

$$\hat{\mathbf{C}} = \begin{bmatrix} \lambda_1 \mathbf{I}_{n\times n} & & & \\ & \lambda_1 \mathbf{I}_{n\times n} & & \\ & & C\mathbf{I}_{m\times m} & \\ & & & C\mathbf{I}_{m\times m} \end{bmatrix} \tag{28a}$$

$$\hat{\mathbf{G}} = \begin{bmatrix} \mathbf{0}_{2n\times 2n} & \mathbf{0}_{2n\times m} & \mathbf{0}_{2n\times m} \\ \mathbf{0}_{m\times 2n} & \mathbf{I}_{m\times m} & \mathbf{0}_{m\times m} \\ \mathbf{0}_{m\times 2n} & \mathbf{0}_{m\times m} & -\mathbf{I}_{m\times m} \end{bmatrix} \tag{28b}$$

$$\hat{\mathbf{A}} = \begin{bmatrix} \mathbf{0}_{2n\times n} & \mathbf{0}_{2n\times n} & \mathbf{0}_{2n\times 2m} \\ -\mathbf{x}_{train} & \mathbf{x}_{train} & \mathbf{0}_{m\times 2m} \\ \mathbf{x}_{train} & -\mathbf{x}_{train} & \mathbf{0}_{m\times 2m} \end{bmatrix} \tag{28c}$$

and the vectors $\mathbf{b}$, $\hat{\mathbf{e}}$, $\hat{\mathbf{d}}$ and $\mathbf{z}$ are defined as:

$$\mathbf{b} = \begin{bmatrix} \mathbf{e}_n \\ \mathbf{e}_n \\ \mathbf{0}_{2m} \end{bmatrix}, \ \hat{\mathbf{e}} = \begin{bmatrix} \mathbf{0}_{2n} \\ \mathbf{e}_m \\ \mathbf{e}_m \end{bmatrix}, \ \hat{\mathbf{d}} = \begin{bmatrix} \mathbf{0}_{2n} \\ \mathbf{y}_{train} \\ -\mathbf{y}_{train} \end{bmatrix}, \ \mathbf{z} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \\ \boldsymbol{\xi} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} \tag{29}$$

The non-negative constraint on the variables $\mathbf{p}$ and $\mathbf{q}$ has been included in Eq. (27b). It can be observed from matric $\hat{\mathbf{C}}$ that there would be zero elements along the diagonal if the linear ε-

insensitive loss function is utilized. Thus, the adoption of quadratic ε-insensitive loss function can enhance the numerical stability in solving optimization problem.

The optimization problem expressed in Eq. (27) can be equivalently solved in the dual formulation. Thus, the Lagrange function $L(\mathbf{z},\gamma,\mathbf{u})$ is shown as Eq. (30):

$$\begin{aligned} L(\mathbf{z},\gamma,\mathbf{u}) = \frac{1}{2}(\mathbf{z}^T\hat{\mathbf{C}}\mathbf{z}+\gamma^2)+\lambda_2\mathbf{b}^T\mathbf{z}-((\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})\mathbf{z} \\ +(\varepsilon\mathbf{I}_{(2m+2n)\times(2m+2n)}+\gamma\hat{\mathbf{G}})\hat{\mathbf{e}}+\hat{\mathbf{d}})^T\mathbf{u} \end{aligned} \tag{30}$$

where $\mathbf{u}\in\Re^{2m+2n}$ denotes the vector contains all Lagrange multipliers. Then, by applying the Karush-Kuhn-Tucker (KKT) conditions for the dual problem, Eq. (30) can be then written as:

$$\begin{aligned} L(\mathbf{u}) &= -\frac{1}{2}\mathbf{u}^T(\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})\hat{\mathbf{C}}^{-1}(\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})^T\mathbf{u}-\frac{1}{2}\lambda_2^2\mathbf{b}^T\hat{\mathbf{C}}^{-1}\mathbf{b} \\ &\quad -\frac{1}{2}\mathbf{u}^T\hat{\mathbf{G}}\hat{\mathbf{e}}\hat{\mathbf{e}}^T\hat{\mathbf{G}}\mathbf{u}+(\lambda_2\mathbf{b}^T\hat{\mathbf{C}}^{-1}(\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})^T-\varepsilon\hat{\mathbf{e}}^T-\hat{\mathbf{d}}^T)\mathbf{u} \\ &= -\frac{1}{2}\mathbf{u}^T((\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})\hat{\mathbf{C}}^{-1}(\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})^T+\hat{\mathbf{G}}\hat{\mathbf{e}}\hat{\mathbf{e}}^T\hat{\mathbf{G}})\mathbf{u} \\ &\quad +(\lambda_2\mathbf{b}^T\hat{\mathbf{C}}^{-1}(\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})^T-\varepsilon\hat{\mathbf{e}}^T-\hat{\mathbf{d}}^T)\mathbf{u}-\frac{1}{2}\lambda_2^2\mathbf{b}^T\hat{\mathbf{C}}^{-1}\mathbf{b} \end{aligned} \tag{31}$$

Since $\lambda_2^2\mathbf{b}^T\hat{\mathbf{C}}^{-1}\mathbf{b}$ is a real constant, the dual problem can be then formulated as:

$$\min_{\mathbf{u}}:\frac{1}{2}\mathbf{u}^T\mathbf{Q}\mathbf{u}-\mathbf{m}^T\mathbf{u} \tag{32a}$$

$$s.t.\quad \mathbf{u}\geq\mathbf{0}_{2m+2n} \tag{32b}$$

where $\mathbf{Q}\in\Re^{(2m+2n)\times(2m+2n)}$ and $\mathbf{m}\in\Re^{2m+2n}$ are defined by:

$$\mathbf{Q}=(\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})\hat{\mathbf{C}}^{-1}(\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})^T+\hat{\mathbf{G}}\hat{\mathbf{e}}\hat{\mathbf{e}}^T\hat{\mathbf{G}} \tag{32c}$$

$$\mathbf{m}^T=\lambda_2\mathbf{b}^T\hat{\mathbf{C}}^{-1}(\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})^T-\varepsilon\hat{\mathbf{e}}^T-\hat{\mathbf{d}}^T \tag{32d}$$

In order to demonstrate that proposed X-SVR has the global minimum solution, we can equivalently prove that the dual problem is a convex optimization problem as the following *Proposition 1*. The details of the proof are demonstrated in Appendix A.

*Proposition 1*:

Given the training dataset with input $\mathbf{x}_{train}\in\Re^{m\times n}$ and output $\mathbf{y}_{train}\in\Re^{m}$, with pre-defining the positive tuning parameters for X-SVR as $\lambda_1,\lambda_2,C,\varepsilon\in\Re^{+}$, the optimization problem defined in Eq. (32) is a convex quadratic programming problem.

Subsequently, the global optimum solution for the proposed X-SVR can be obtained by solving the associated dual problem by either gradient based method or available quadratic programming

solvers. Let $\mathbf{u}^* \in \Re^{2m+2n}$ be the obtained solution for the X-SVR, then the variables $\mathbf{z}$ and $\gamma$ can be respectively computed as:

$$\mathbf{z} = \hat{\mathbf{C}}^{-1}((\hat{\mathbf{A}} + \mathbf{I}_{(2m+2n)\times(2m+2n)})^T \mathbf{u}^* - \lambda_2 \mathbf{b}) \tag{33}$$

$$\gamma = \hat{\mathbf{e}}^T \hat{\mathbf{G}} \mathbf{u}^* \tag{34}$$

Then, the coefficient $\mathbf{w}$ can be obtained as:

$$\mathbf{w} = \mathbf{p} - \mathbf{q} = \mathbf{z}(1:n) - \mathbf{z}(n+1:2n) \tag{35}$$

Thus, the regression function obtained by the proposed by the proposed X-SVR is expressed as:

$$\hat{f}(\mathbf{x}) = (\mathbf{p} - \mathbf{q})^T \mathbf{x} - \hat{\mathbf{e}}^T \hat{\mathbf{G}} \mathbf{u}^* \tag{36}$$

Similar as the classic support vector regression, the proposed X-SVR can be further modified into a nonlinear regression method (using kernel method) such that the introduced approach can be applied on the more complex problems. Despite of the convenience of the kernel method used in the ε-SVR, mapping to the intrinsic vector space can only applied as a replacement of $\mathbf{x}_i^T \mathbf{x}_j$ in order to avoiding the explicit calculation of $\mathbf{\Phi}(\mathbf{x}_i)$ and $\mathbf{\Phi}(\mathbf{x}_j)$. In the proposed X-SVR, such implicit kernel map approach is not applicable since that the linear combinations of $\mathbf{x}_i$ ($i = 1, 2, ..., m$), in addition to the inner product $\mathbf{x}_i^T \mathbf{x}_j$ are included in the dual formulation expressed as Eq. (32) To extend the linear X-SVR to a kernelized learning method, an alternative approach, namely the *empirical kernel map* [43, 51], is utilized in the proposed surrogate model. The adopted empirical kernelization can be expressed in Eq. (38):

$$\mathbf{x}_i = [x_{i,1}, x_{i,2}, ..., x_{i,n}]^T \mapsto \hat{\mathbf{k}}(\mathbf{x}_i) = \begin{bmatrix} \mathbf{\Phi}(\mathbf{x}_1)^T \mathbf{\Phi}(\mathbf{x}_i) \\ \mathbf{\Phi}(\mathbf{x}_2)^T \mathbf{\Phi}(\mathbf{x}_i) \\ \vdots \\ \mathbf{\Phi}(\mathbf{x}_m)^T \mathbf{\Phi}(\mathbf{x}_i) \end{bmatrix} = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_i) \\ K(\mathbf{x}_2, \mathbf{x}_i) \\ \vdots \\ K(\mathbf{x}_m, \mathbf{x}_i) \end{bmatrix} \tag{37}$$

where the kernel-induced vector $\hat{\mathbf{k}}(\mathbf{x}_i)$ is known as the *empirical feature vector* with the *empirical degree* $m$ defined as the number of training samples [43]. Such $m$-dimensional vector space is named as the *empirical feature space*. Then, the empirical feature vector $\hat{\mathbf{k}}(\mathbf{x}_i)$ is regarded as the training sample for constructing the learning model. Comparing with the implicit kernel map approach used in ε-SVR, the empirical feature space is finite-dimensional and jointly defined by the employed kernel function and training samples [51]. Such kernel map approach has also been effectively applied on the other kernelized learning method, including kernelized LASSO (Least

Absolute Selection and Shrinkage Operator) [52], kernelized elastic net [53] and linear programming SVR [54]. The architecture of the nonlinear X-SVR is shown in Fig. 3.

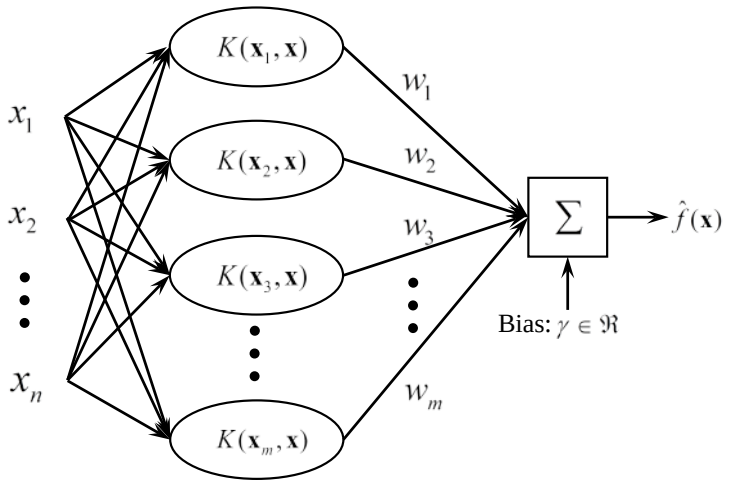



Fig. 3 The architecture of the X-SVR

Thus, given the training dataset $\mathbf{x}_{train}$ and a specific kernel function $K(\mathbf{x}_i,\mathbf{x}_j)$, the original training samples are transferred into the kernel matrix $\mathbf{K}_{train} \in \Re^{m\times m}$ which is expressed as Eq. (38):

$$\mathbf{K}_{train} = \begin{bmatrix} K(\mathbf{x}_1,\mathbf{x}_1) & K(\mathbf{x}_1,\mathbf{x}_2) & \cdots & K(\mathbf{x}_1,\mathbf{x}_m) \\ K(\mathbf{x}_2,\mathbf{x}_1) & K(\mathbf{x}_2,\mathbf{x}_2) & \cdots & K(\mathbf{x}_2,\mathbf{x}_m) \\ \vdots & \vdots & \cdots & \vdots \\ K(\mathbf{x}_m,\mathbf{x}_1) & K(\mathbf{x}_m,\mathbf{x}_2) & \cdots & K(\mathbf{x}_m,\mathbf{x}_m) \end{bmatrix} \tag{38}$$

Then, the kernel matrix $\mathbf{K}_{train}$ is used as the training dataset and the nonlinear X-SVR problem is now formulated as Eq. (39):

$$\min_{\mathbf{p}_k,\mathbf{q}_k,\gamma,\xi,\hat{\xi}} : \frac{\lambda_1}{2}(\|\mathbf{p}_k\|_2^2 + \|\mathbf{q}_k\|_2^2) + \lambda_2 \mathbf{e}_m^T(\mathbf{p}_k + \mathbf{q}_k) + \frac{C}{2}(\xi^T\xi + \hat{\xi}^T\hat{\xi}) \tag{39a}$$

$$s.t. \quad \begin{cases} \mathbf{K}_{train}(\mathbf{p}_k - \mathbf{q}_k) - \gamma\mathbf{e}_m^T - \mathbf{y}_{train} \le \varepsilon\mathbf{e}_m^T + \xi \\ \mathbf{y}_{train} - (\mathbf{K}_{train}(\mathbf{p}_k - \mathbf{q}_k) - \gamma\mathbf{e}_m^T) \le \varepsilon\mathbf{e}_m^T + \hat{\xi} \\ \mathbf{p}_k, \mathbf{q}_k, \xi, \hat{\xi} \ge \mathbf{0}_m \end{cases} \tag{39b}$$

where $\mathbf{p}_k, \mathbf{q}_k \in \Re^m$ and have the same function as $\mathbf{p}$ and $\mathbf{q}$ for linear X-SVR; the subscript $k$ is for indicating that this is a kernelized learning model. Then, by adopting the same concept as expressed in Eq. (27), the kernelized X-SVR shown in Eq. (39) can be modified into:

$$\min_{\mathbf{z}_k,\gamma} : \frac{1}{2}(\mathbf{z}_k^T\hat{\mathbf{C}}\mathbf{z}_k + \gamma^2) + \lambda_2\mathbf{b}_k^T\mathbf{z}_k \tag{40a}$$

$$s.t. \quad (\hat{\mathbf{A}}_k + \mathbf{I}_{4m\times 4m})\mathbf{z}_k + (\varepsilon\mathbf{I}_{4m\times 4m} + \gamma\hat{\mathbf{G}}_k)\hat{\mathbf{e}}_k + \hat{\mathbf{d}}_k \ge \mathbf{0}_{4m} \tag{40b}$$

where $\mathbf{I}_{4m\times 4m} \in \Re^{4m\times 4m}$ denotes the identity matrix and $\mathbf{0}_{4m} \in \Re^{4m}$ denotes the zero matrix. The matrices $\hat{\mathbf{C}}_k$, $\hat{\mathbf{G}}_k$ and $\hat{\mathbf{A}}_k$ are defined as:

$$\hat{\mathbf{C}}_k = \begin{bmatrix} \lambda_1\mathbf{I}_{m\times m} & & & \\ & \lambda_1\mathbf{I}_{m\times m} & & \\ & & C\mathbf{I}_{m\times m} & \\ & & & C\mathbf{I}_{m\times m} \end{bmatrix} \tag{41a}$$

$$\hat{\mathbf{G}}_k = \begin{bmatrix} \mathbf{0}_{2m\times 2m} & \mathbf{0}_{2m\times m} & \mathbf{0}_{2m\times m} \\ \mathbf{0}_{m\times 2m} & \mathbf{I}_{m\times m} & \mathbf{0}_{m\times m} \\ \mathbf{0}_{m\times 2m} & \mathbf{0}_{m\times m} & -\mathbf{I}_{m\times m} \end{bmatrix} \tag{41b}$$

$$\hat{\mathbf{A}}_k = \begin{bmatrix} \mathbf{0}_{2m\times m} & \mathbf{0}_{2m\times m} & \mathbf{0}_{2m\times 2m} \\ -\mathbf{K}_{train} & \mathbf{K}_{train} & \mathbf{0}_{m\times 2m} \\ \mathbf{K}_{train} & -\mathbf{K}_{train} & \mathbf{0}_{m\times 2m} \end{bmatrix} \tag{41c}$$

and the vectors $\mathbf{b}_k$, $\hat{\mathbf{e}}_k$, $\hat{\mathbf{d}}_k$ and $\mathbf{z}_k$ are defined as:

$$\mathbf{b}_k = \begin{bmatrix} \mathbf{e}_m \\ \mathbf{e}_m \\ \mathbf{0}_{2m} \end{bmatrix},\ \hat{\mathbf{e}}_k = \begin{bmatrix} \mathbf{0}_{2m} \\ \mathbf{e}_m \\ \mathbf{e}_m \end{bmatrix},\ \hat{\mathbf{d}}_k = \begin{bmatrix} \mathbf{0}_{2m} \\ \mathbf{y}_{train} \\ -\mathbf{y}_{train} \end{bmatrix},\ \mathbf{z}_k = \begin{bmatrix} \mathbf{p}_k \\ \mathbf{q}_k \\ \boldsymbol{\xi} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} \tag{42}$$

The optimization problem demonstrated in Eq. (40) can also be equivalently solved in its dual formulation by using Lagrange method with KKT conditions. Thus, by introducing the non-negative Lagrange multiplier $\mathbf{u}_k \in \Re^{4m}$, the kernelized X-SVR will be solved as the quadratic programming problem shown in Eq. (43):

$$\min_{\mathbf{u}_k} : \frac{1}{2}\mathbf{u}_k^T\mathbf{Q}_k\mathbf{u}_k - \mathbf{m}_k^T\mathbf{u}_k \tag{43a}$$

$$s.t. \quad \mathbf{u}_k \geq \mathbf{0}_{4m} \tag{43b}$$

where $\mathbf{Q}_k \in \Re^{4m\times 4m}$ and $\mathbf{m}_k \in \Re^{4m}$ are defined by:

$$\mathbf{Q}_k = (\hat{\mathbf{A}}_k + \mathbf{I}_{4m\times 4m})\hat{\mathbf{C}}_k^{-1}(\hat{\mathbf{A}}_k + \mathbf{I}_{4m\times 4m})^T + \hat{\mathbf{G}}_k\hat{\mathbf{e}}_k\hat{\mathbf{e}}_k^T\hat{\mathbf{G}}_k \tag{43c}$$

$$\mathbf{m}_k^T = \lambda_2\mathbf{b}_k^T\hat{\mathbf{C}}_k^{-1}(\hat{\mathbf{A}}_k + \mathbf{I}_{4m\times 4m})^T - \varepsilon\hat{\mathbf{e}}_k^T - \hat{\mathbf{d}}_k^T \tag{43d}$$

Let $\mathbf{u}_k^* \in \Re^{4m}$ be the obtained solution for the optimization problem illustrated in Eq. (43), then the variables $\mathbf{z}_k$ and $\gamma_k$ can be respectively computed as:

$$\mathbf{z}_k = \hat{\mathbf{C}}_k^{-1}((\hat{\mathbf{A}}_k + \mathbf{I}_{4m\times 4m})^T\mathbf{u}_k^* - \lambda_2\mathbf{b}_k) \tag{44}$$

$$\gamma = \hat{\mathbf{e}}_k^T \hat{\mathbf{G}}_k \mathbf{u}_k^* \tag{45}$$

Then, the coefficient $\mathbf{w}$ can be obtained as:

$$\mathbf{w} = \mathbf{p}_k - \mathbf{q}_k = \mathbf{z}_k(1:m) - \mathbf{z}_k(m+1:2m) \tag{46}$$

Thus, the nonlinear regression function obtained by the proposed kernelized X-SVR is expressed as:

$$\hat{f}(\mathbf{x}) = (\mathbf{p}_k - \mathbf{q}_k)^T \hat{\mathbf{k}}(\mathbf{x}) - \hat{\mathbf{e}}_k^T \hat{\mathbf{G}}_k \mathbf{u}_k^* \tag{47}$$

It can be easily observed that the only difference between the linear and nonlinear X-SVR is that the input dataset is mapped into the empirical space by using specified kernel function. Thus, the kernelized X-SVR can be regarded as a linear X-SVR with a manipulated input samples and therefore the convex property is still promised regardless of the type of kernel function. In this context, the selected kernel function is not restricted to the category that satisfies the Mercer's theorem [53].

### *3.3 Generalized Gegenbauer kernel – a new orthogonal polynomial kernel function*

For both nonlinear classification and regression applications, the performance of support vector machine is significantly affected by the employed kernel functions [55]. Despite that Gaussian and polynomial kernels are commonly adopted, it is pointed out that these kernels can lead to unsatisfied results in approximating some complex function [54, 55]. Specifically, Gaussian and polynomial kernels are not complete orthonormal base, which result in that they cannot approach to the curves in quadratic continuous integral space [56]. To overcome such drawback, the wavelet kernel function is proposed and receives favorable results for both classification and regression [54-56]. Motivated by the properties of orthogonal polynomials which have been effectively used for functions approximation, the development of orthogonal polynomial kernels for SVM/ SVR models receives increasingly attention from researchers [57-60].

Among the family of orthogonal polynomials, the Gegenbauer polynomial has been widely adopted for uncertainty quantification and function approximation by using the Gegenbauer series expansion [61]. The univariate Gegenbauer polynomials, denoted by $P_d^{\alpha_k}(x)$, with degree $d \in \mathbb{Z}$ and polynomial parameter $\alpha_k > 0$ can be defined by the recurrence relations as Eq. (48):

$$\begin{cases} P_0^{\alpha_k}(x) = 1 \\ P_1^{\alpha_k}(x) = 2\alpha_k x \\ P_d^{\alpha_k}(x) = \dfrac{1}{d}[2x(d+\alpha_k-1)P_{d-1}^{\alpha_k}(x) - (d+2\alpha_k-2)P_{d-2}^{\alpha_k}(x)], \quad d = 2,3,4,\ldots \end{cases} \tag{48}$$

For a given $\alpha_k$, the Gegenbauer polynomials are orthogonal on $x \in [-1,1]$ with respect to the weight function $\rho^{\alpha_k}(x)$, which can be expressed as

$$\int_{-1}^{1} \rho^{\alpha_k}(x) P_l^{\alpha_k}(x) P_v^{\alpha_k}(x) dx = \begin{cases} h_l^{\alpha_k}, & l = v \\ 0, & l \neq v \end{cases} \tag{49}$$

where $l, v = 0,1,2,...,d$ and $\rho^{\alpha_k}(x)$ and $h_l^{\alpha_k}$ can be formulated as

$$\rho^{\alpha_k}(x) = (1-x^2)^{\alpha_k - \frac{1}{2}} \tag{50a}$$

$$h_l^{\alpha_k} = \frac{\pi 2^{1-2\alpha_k} \Gamma(d+2\alpha_k)}{d!(d+\alpha_k)\Gamma^2(\alpha_k)} \tag{50b}$$

In Eq. (50b), $\Gamma(\bullet)$ denotes the Gamma function. As the particular solutions of the Gegenbauer differential equation, such polynomial is the generalization of Chebyshev and Legendre polynomials by substituting various value for $\alpha_k$ [62].

Considering generalization ability of Gegenbauer polynomial, Padierna et al. [63] proposed a new orthogonal polynomial kernel based on the Gegenbauer polynomial and implemented on binary classification problems. Similar as the Legendre and Hermite polynomial kernels, the Gegenbauer polynomial kernel is constructed as the tensor product of the inner product of univariate polynomials, which is conceptually identical with the method for extending one-dimensional polynomials to multi-dimensional. As pointed out by [57], the kernel constructed by the tensor product approach may yield either an extremely small and larger value which will significantly impact the performance the corresponding kernelized learning models. Such phenomenon is avoided in the Gegenbauer polynomial kernel by multiplying weight and scaling functions to the product univariate polynomials and limiting the variation range of the polynomial parameter [63] However, it is point out by Ozer et al. [57] that such type of kernel construction approach may force the learning along each input variable rather than the input vectors. Thus, it is suggested that the orthogonal polynomial kernel functions should be applied directly onto the input vectors rather than each pair of input elements.

Inspired by the pioneering work by Ozer et al [57] and Padierna et al. [63], we developed a new orthogonal polynomial kernel function for the proposed kernelized X-SVR. Different from the kernel function introduced by [63], the proposed orthogonal polynomial kernel is constructed by using the partial sum of the inner product of generalized Gegenbauer polynomials, namely the *generalized Gegenbauer kernel* (GGK). By adopting the strategy utilized for defining the generalized Chebyshev polynomial for vector inputs [57], the generalized Gegenbauer polynomials are defined recursively as following:

$$\begin{cases} P_0^{\alpha_k}(\mathbf{x}) = 1 \\ P_1^{\alpha_k}(\mathbf{x}) = 2\alpha_k \mathbf{x} \\ P_d^{\alpha_k}(\mathbf{x}) = \dfrac{1}{d}[2\mathbf{x}^T(d+\alpha_k-1)P_{d-1}^{\alpha_k}(\mathbf{x}) - (d+2\alpha_k-2)P_{d-2}^{\alpha_k}(\mathbf{x})], \quad d = 2,3,4,... \end{cases} \tag{51}$$

where $\mathbf{x} \in \Re^n$ denotes the column vector of input variables. It can be revealed from Eq. (51) that the generalized Gegenbauer polynomial $P_d^{\alpha_k}(\mathbf{x})$ yields a scalar value when the polynomial order $d$ is an even number, otherwise it will yield a column vector. Considering that an exponential function, such as Gaussian kernel function, has better capability in capturing local information than the originally employed square root function [57], the Gaussian kernel function is adopted here as the weighting function for the proposed GGK. Thus, the proposed *n*th order generalized Gegenbauer Kernel function $K_{GGK}(\mathbf{x}_i, \mathbf{x}_j)$ of two arbitrary input vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined as Eq. (52):

$$K_{GGK}(\mathbf{x}_i, \mathbf{x}_j) = \frac{\sum_{l=0}^{d} P_l^{\alpha_k}(\mathbf{x}_i)^T P_l^{\alpha_k}(\mathbf{x}_j)}{\exp(\sigma \left\| \mathbf{x}_i - \mathbf{x}_j \right\|_2^2)} \tag{52}$$

where each element of $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined in $[-1,1]$. In this context, both $\alpha_k$ and $\sigma$ are here considered as the kernel scales or the so-called decaying parameters of the proposed kernel function.

It is worthy to addressing that the proposed GGK satisfies the Mercer Theorem which is a prerequisite for implementing the kernel function in SVM/SVR. Thus, not just in the proposed X-SVR model, the generalized Gegenbauer kernel introduced in this study can be also employed in the other kernelized learning models which require the Mercer condition to be satisfied. The property that the proposed GGK is a valid Mercer kernel can be systematically demonstrated by the *Proposition 2*. The proof of the *Proposition 2* is demonstrated in the Appendix B. It can be observed from Eq. (52), the novel GGK possesses three kernel parameters: the polynomial order $d$ and two positive kernel scale parameters $\alpha_k$ and $\sigma$.

*Proposition 2*:

The proposed generalized Gegenbauer kernel (GGK) expressed in Eq. (52) is a valid Mercer kernel.

It can be observed from the proof of *Proposition 2* that the proposed GGK is a mixed kernel function which combines one local kernel $K_1(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\sigma \left\| \mathbf{x}_i - \mathbf{x}_j \right\|_2^2)$ (Gaussian kernel) and one global kernel $K_2(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=0}^{d} P_l^{\alpha_k}(\mathbf{x}_i)^T P_l^{\alpha_k}(\mathbf{x}_j)$ (a generalized polynomial kernel) [58]. Subsequently,

by integrating the proposed generalized Gegenbauer kernel, the nonlinear X-SVR can be also regarded as a multiple kernel learning algorithm using fixed rules approach [64].

### *3.4 Hyperparameter optimization for X-SVR with generalized Gegenbauer kernel*

In the proposed X-SVR with the generalized Gegenbauer kernel (GGK), there are seven hyperparameters including the two regularization parameters $\lambda_1$ and $\lambda_2$, the penalty parameter $C$, the insensitive tube width $\varepsilon$, the polynomial order $d$ and two positive kernel scale parameters $\alpha_k$ and $\sigma$. Similar as the conventional SVR model, the prediction accuracy of the proposed X-SVR with GGK is strongly dependent on the selection of these parameters. For machine learning approaches, the $k$-fold cross-validation (CV) over the training samples is an effective approach to ensure the regression model has the generalized ability in accurately predicting the training dataset while checking if the selected parameters will result in overfitting [65]. Practically, $k$ is commonly set to 5-10 as a trade-off of computational cost and prediction accuracy [66]. In present work, the 5-fold CV error which denoted by $Err_{5CV}$ is employed as the training error measure for X-SVR, which is formulated as following:

$$Err_{5CV} = \frac{1}{5}\sum_{i=1}^{5} err_i \tag{53}$$

where $err_i$ is the mean squared error (MSE) between the predicted output $\hat{f}(\mathbf{x})$ obtained by the X-SVR model and the output of the true function $f(\mathbf{x})$ in each fold $i$. $err_i$ is expressed as Eq. (54):

$$err_i = \frac{1}{m_i}\sum_{j=1}^{m_i}(y_{i,j} - \hat{f}(\mathbf{x}_{i,j}))^2 \tag{54}$$

where $m_i$ denotes the number of training samples in the fold $i$; $\mathbf{x}_{i,j}$ and $y_{i,j}$ denote the $j$th input and output in the $i$th fold, respectively.

Table 1 Searching range of X-SVR hyperparameters

| Hyperparameter | Searching range |
|---|---|
| $\lambda_1$ | $[10^{-3},10^{3}]$ |
| $\lambda_2$ | $[10^{-3},10^{3}]$ |
| $C$ | $[10^{1},10^{5}]$ |
| $\varepsilon$ | $[10^{-5},10^{-1}]$ |
| $d$ | $[1,6]$ |
| $\alpha_k$ | $[10^{-3},1]$ |

| | |
|---|---|
| $\sigma$ | $[10^{-2},10^{2}]$ |

Since that the selected hyperparameters will lead to the minimization of $Err_{5CV}$ , the hyperparameter tuning can be considered as an optimization problem. Recently, Bayesian optimization is becoming increasingly popular in tuning learning parameters for complex machine learning algorithm such as deep neural network [67]. Typically, Bayesian optimization construct a probabilistic approximation of the objective function by using Gaussian process and then determines the next estimation point which results in the maximum of the acquisition function [68]. Instead of using the local gradient or Hessian approximations, Bayesian optimization relies on all the available information from previous evaluations of the objective function. Subsequently, the minimum of the objective function can be efficiently obtained with relative less number of iteration [68]. Considering that more hyperparameters are included in proposed X-SVR model with generalized Gegenbauer kernel in comparison with the classic nonlinear ε-SVR, Bayesian optimization method is integrated in the proposed meta-model for automatically selecting the suitable learning parameters. In the presented work, the Bayesian optimization is conducted by using the MATLAB Statistical and Machine Leaning toolbox [69]. The searching range for the hyperparameters are summarized in Table 1.

## 4. Structural dynamic reliability analysis by using X-SVR

This paper offers a metamodel-based Monte Carlo Simulation method for structural dynamic reliability analysis by adopting the proposed X-SVR with the generalized Gegenbauer kernel. In this proposed reliability analysis strategy, the true structural limit state function is approximated by using the X-SVR metamodel to replace the precise FEM model. Then, the probability of failure is evaluated by conducting the Monte Carlo simulation based on the constructed surrogate model. Thus, the performance of the proposed method is significantly affected by the quality of the trained X-SVR model. In practice, metamodels are trained by limit number of running the original models based on the Design of Experiments (DoEs) and expected to have good predictions over the entire domain of input variables, which is commonly achieved by employing uniform sampling techniques [70, 71]. In this study, the training samples are generated by quasi-Monte Carlo sampling method with Sobol's sequence [72]. Such low-discrepancy sampling technique can generate samples evenly distributed over the design space. Then, the X-SVR surrogate model will be trained by using the DoE and subsequently used for analyzing the probability of failure ( $p_f$ ).

In the presented study, the first passage problem with equal barriers ( $b_1 = -b_2 = b$ ) is investigated for demonstrating the capability of the proposed metamodel-based reliability analysis approach. The procedure for this reliability analysis is summarized as follows:

1. Generate $m_{MCS}$ Monte Carlo samples $\mathbf{x}_{MCS} \in \Re^{m_{MCS} \times n}$ for input variables using the quasi-MCS scheme. The generated samples will be used as the input samples for the trained metamodel for reliability analysis using metamodel. $m_{MCS}$ is expected to be large (i.e. $10^5 - 10^6$).
2. Define the DoE with $m_{train}$ ( $m_{train} \ll m_{MCS}$ ) training samples $\mathbf{x}_{train}$ and $\mathbf{y}_{train}$ using the quasi-MCS (Sobol's sequence). Here $y_{train,i} = \max \left| u(\mathbf{x}_i, t) \right|$ for $i = 1, ..., m_{train}$, which is obtained by high-fidelity numerical analysis (i.e. FEM) in this study.
3. Train the X-SVR model based on the samples $\{\mathbf{x}_{train}, \mathbf{y}_{train}\}$ obtained in step 2. The limit state function of the investigated structure can be approximately expressed as $\hat{g}(\mathbf{x})$ by employing the trained X-SVR surrogate model.
4. Input the MCS samples into the metamodel-based limit state function $\hat{g}(\mathbf{x})$ to analysis the dynamic response. Compute the number of failure samples $m_f$ which are indicated by $\hat{g}(\mathbf{x}) \le 0$.
5. The probability of failure for the investigated structure is approximated calculated as $\hat{p}_f = m_f / m_{MCS}$.

The flowchart of the computational procedures is shown in Fig. 4.

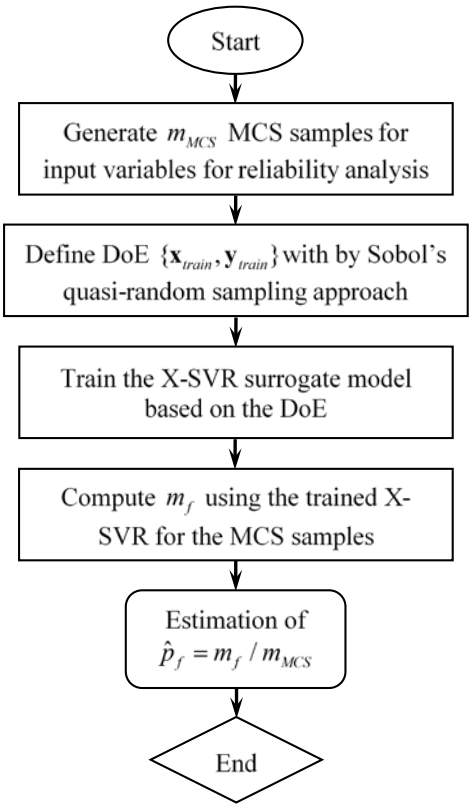



Fig. 4. Flowchart for the X-SVR based structural dynamic reliability analysis

## 5. Numerical examples

To demonstrate the capability and accuracy of the proposed approach, four numerical examples are presented in this study. The first two analytical examples are benchmark functions which are adopted here for illustrating the performance of the proposed approach. Then, the proposed approach is further tested by one structural dynamic problem and one acoustic problem for demonstrating its reliability and versatility. The results are compared with the classical ε – support vector regression (ε-SVR) model with widely used Gaussian kernel. Moreover, the direct Monte-Carlo simulation (MCS) is conducted for each example as a reference for comparing the accuracy of the two methods. In all the considered examples, both the proposed method and ε-SVR model are trained by samples generated by Quasi MCS with Sobol sequence [72]. In this work, the ε-SVR integrated in the MATLAB Statistical and Machine Leaning toolbox [69] is adopted. The presented numerical results are obtained by using a workstation with CPU of Intel Core i7-4770, 32 GB of memory, and 1 TB of hard drive.

### *5.1 First analytical example: Borehole function*

The first example employed for demonstrating the performance of the proposed method is called the Borehole function which is commonly used as a benchmark example for emulation and prediction tests [73]. This function, as expressed in Eq. (58), was originally derived for modelling the

water flow through a borehole. There are totally 8 input parameters which are all modeled as independent and uniformly distributed variables. The details of the variation range are presented in Table 2.

$$f(\mathbf{x}) = \frac{2\pi x_3 (x_4 - x_6)}{\ln(x_2 / x_1)(1 + \frac{2x_7 x_3}{\ln(x_2 / x_1) x_1^2 x_8} + \frac{x_3}{x_5})} \tag{55}$$

Table 2 Details of the input parameters for the Borehole function

| Input parameters | Range |
|---|---|
| $x_1$ | $[0.05, 0.15]$ |
| $x_2$ | $[100, 50000]$ |
| $x_3$ | $[63070, 115600]$ |
| $x_4$ | $[990, 1110]$ |
| $x_5$ | $[63.1, 116]$ |
| $x_6$ | $[700, 820]$ |
| $x_7$ | $[1120, 1680]$ |
| $x_8$ | $[9855, 12045]$ |

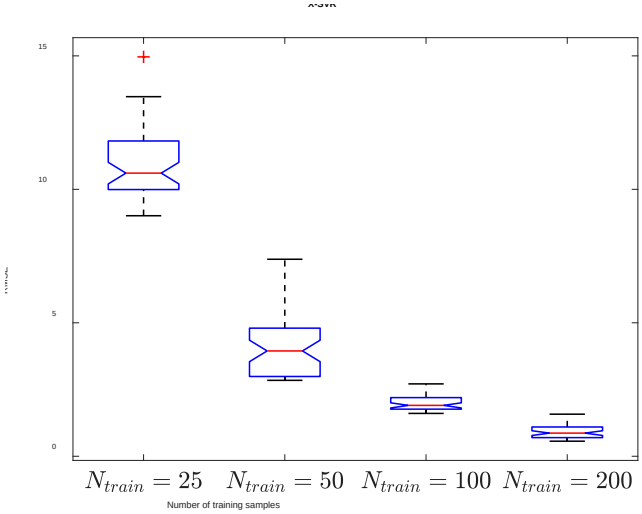
The performance of the proposed model is tested with a variety number of training samples ($N_{train}$). In this example, the initial design of experiment (DoE) consists 25 sampling points and then augmented to 50, 100 until 200 samples. Moreover, the classical ε-SVR model with Gaussian kernel is applied with the same DoEs for comparison purpose. The accuracy of the metamodels is assessed by evaluating two types of relative error: the root mean squared error (RMSE) and the coefficient of determination ($R^2$) which are described as in Table 3, where $f(\mathbf{x}_i)$ denotes the output of the actual model at the sampling point $\mathbf{x}_i$; $\hat{f}(\mathbf{x}_i)$ denotes the output of the constructed surrogate model at the sampling point $\mathbf{x}_i$; and $\mu_f$ denotes the estimated mean of the outputs of all the $N_{MCS}$ sampling points for the actual model. The RMSE is scale-dependent to the magnitude of data to be predicted and a lower RMSE value indicates a higher accuracy of the surrogate model. The $R^2$ offers a statistical measure of the goodness of regression predictions in approximating the real data points. Surrogate models are indicated to have better capability in prediction if $R^2$ is closer to 1. The two validation

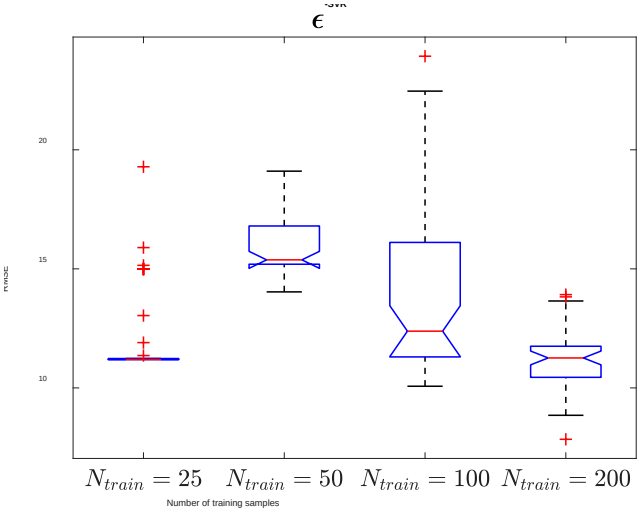
errors are computed for each surrogate model with different numbers of training samples by using the MCS with $N_{MCS} = 5 \times 10^4$ samples in this example.

Table 3 The expressions of RMSE and $R^2$

| Error metrics | Expression |
|---|---|
| Root mean square error (RMSE) | $RMSE = \sqrt{\frac{1}{N_{MCS}} \sum_{i=1}^{N_{MCS}} (f(\mathbf{x}_i) - \hat{f}(\mathbf{x}_i))^2}$ |
| Coefficient of determination ( $R^2$ ) | $R^2 = 1 - \frac{\sum_{i=1}^{N_{MCS}} (f(\mathbf{x}_i) - \hat{f}(\mathbf{x}_i))^2}{\sum_{i=1}^{N_{MCS}} (f(\mathbf{x}_i) - \mu_f)^2}$ |

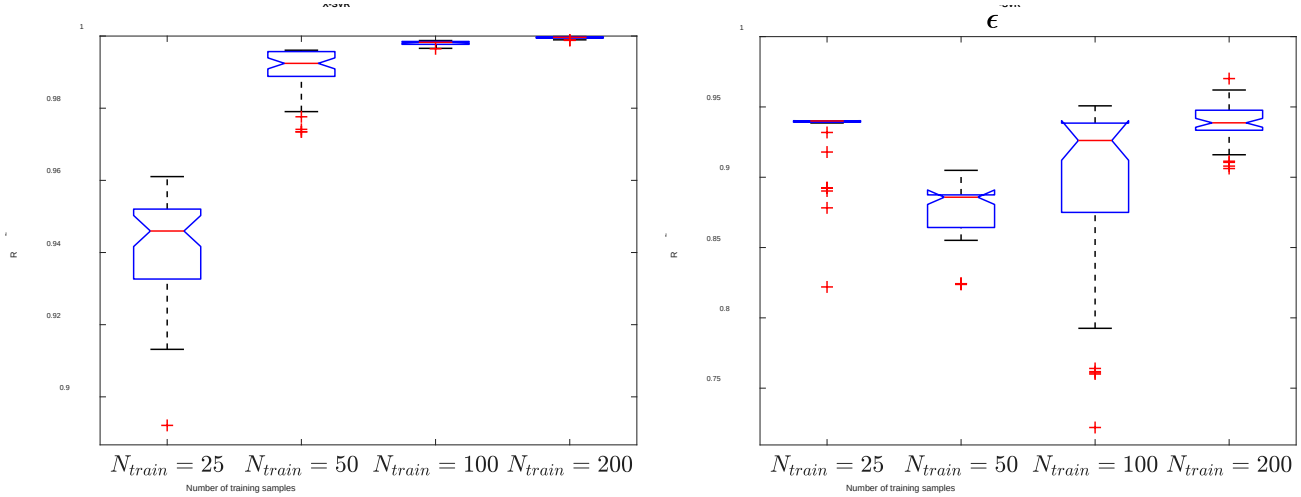
To statistically assess the predicting performance of the surrogate models, the analysis is replicated 50 times at each DoE for both methods and boxplots of the RMSE and $R^2$ of predicted results are shown in Figs. 5 and 6, respectively. In the figures, the median error, the quantile error values and the extreme error values of the 50 independently repeated simulations are demonstrated. As shown in Fig. 5, the proposed X-SVR method has a better predicting performance than the conventional ε-SVR in terms of median value. It is also noticed from Fig. 5 that, for $N_{train} = 25$, more outliers are shown in the boxplot of RMSE of the simulation results obtained by ε-SVR than the proposed approach. Additionally, both the median value and variation range for the RMSE of the X-SVR prediction decrease with the increase of training sample size. This phenomenon is not clearly demonstrated from Fig. 5 for the ε-SVR. Similarly, it is shown in Fig.6 that the proposed surrogate model has the less scattered $R^2$ value than the conventional SVR method. Thus, it can be concluded that the proposed model has a better performance in this example.



(a) (b)

Fig. 5. The boxplots of RMSE for 50 independent simulations of (a) X-SVR and (b) ε-SVR trained by different DoEs




(a) (b)

Fig. 6. The boxplots of $R^2$ for 50 independent simulations of (a) X-SVR and (b) ε-SVR trained by different DoEs

*5.2 Second analytical example: 50-D function*

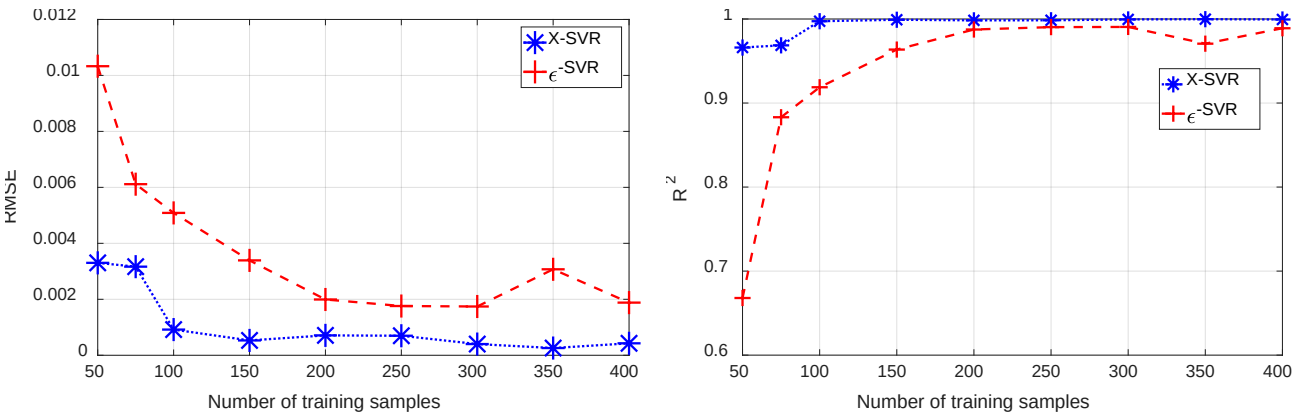
For testing the capability of the proposed X-SVR method for high-dimensional problems, an analytical example with 50 input parameters is utilized. Such 50-D function [74] has been widely used in evaluating the performance of optimization algorithm in high dimensional space. The considered function is expressed in Eq. (56). In this example, the input variables are assumed to be independent and uniformly distributed within the range $[0,1]$.

$$f(\mathbf{x}) = 1 - \exp(-0.01\sum_{i=1}^{50} x_i^2); x_i \in [0,1] \tag{56}$$

Similar as in the Example 1, the performance of the proposed X-SVR with generalized Gegenbauer kernel is investigated by comparing the ε-SVR with widely used Gaussian kernel against various numbers of training samples ($N_{train}$). Considering of the relatively large number of input variables, the initial DOE is selected as 50 and increased gradually to 400. The validation errors RMSE and $R^2$ for both surrogate models trained with different $N_{train}$ are computed using $N_{MCS} = 5\times10^4$ MCS samples and are plotted in Fig. 7. It can be observed from simulation results that, the RMSE for X-SVR is less than that for ε-SVR while the $R^2$ for X-SVR is larger than that for ε-SVR with any adopted $N_{train}$, which indicates that the X-SVR with generalized Gegenbauer kernel

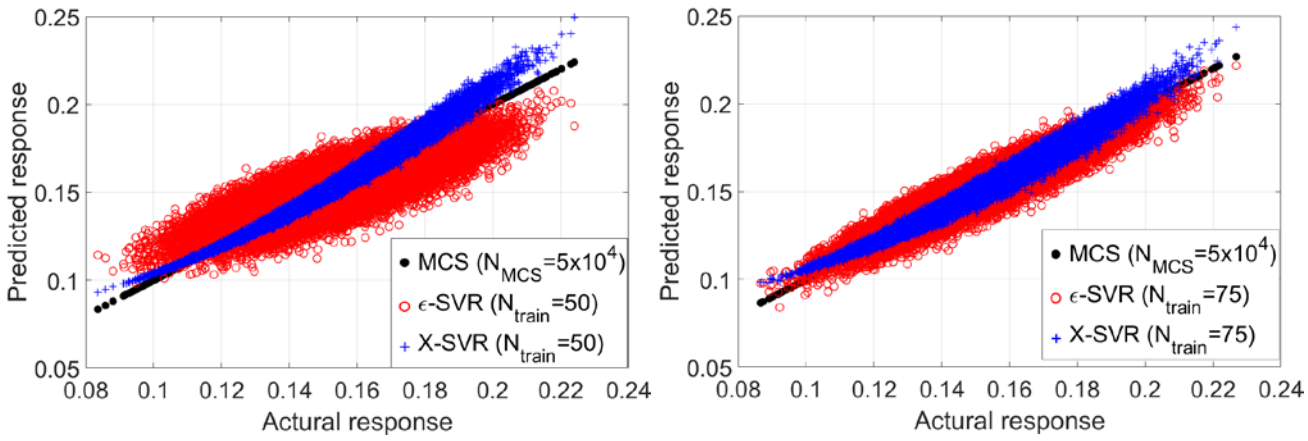
outperforms the ε-SVR with Gaussian kernel in this example. Despite of that the validation errors reduce with the increase of $N_{train}$ for both methods, the RMSE and $R^2$ for the proposed ε-SVR converge faster than that for the classic ε-SVR. Under the circumstance of small number of training samples ( $N_{train} \le 150$ ), it is shown in Fig. 7(a) that the X-SVR has much less RMSE than the ε-SVR.

In order to offer a visible demonstration, the scatter plots of prediction results (predicted response in Fig. 8) obtained by both X-SVR and ε-SVR trained with relatively small number of DOEs are shown in Fig. 8. It can be visualized that, with $N_{train} = 50$ and $75$, the prediction obtained by X-SVR is less variant from the true value (actual response in Fig. 8) of the 50D function obtained by direct MCS. In the context that $N_{train} = 100$ and $150$, the function value predicted by X-SVR is almost identical to the MCS results by observation. Thus, the proposed X-SVR with generalized Gegenbauer kernel has superior capability in approximating the 50D function than the ε-SVR with Gaussian kernel.

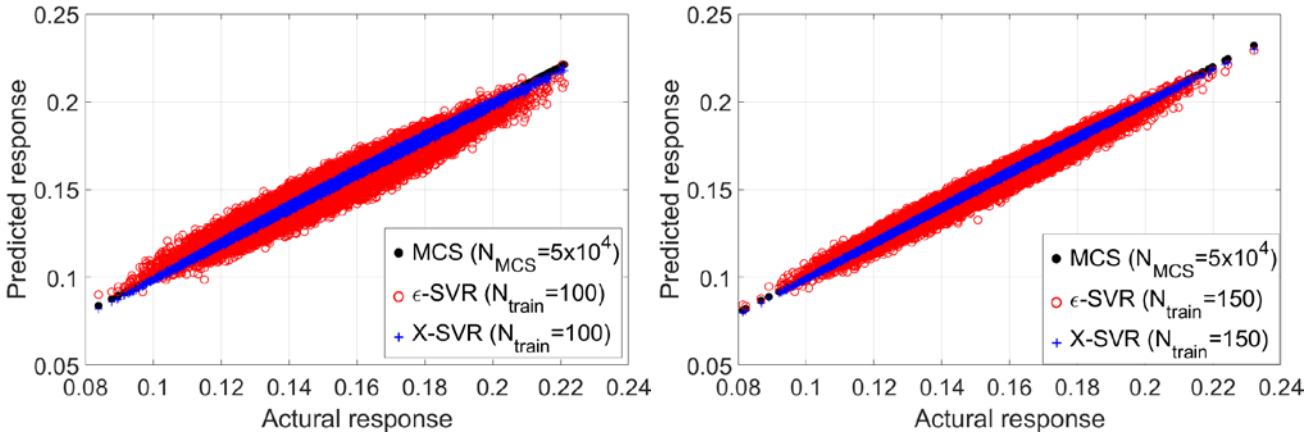



(a) (b)

Fig. 7. The (a) RMSE and (b) $R^2$ of X-SVR and ε-SVR trained with various DOEs




(a) (b)




(c) (d)

Fig. 8. Comparative assessment of by X-SVR and ε-SVR utilizing various number of training samples: (a) $N_{train}=50$, (b) $N_{train}=75$, (c) $N_{train}=100$ and (d) $N_{train}=150$

*5.3 Spring-mass-damper system with three degrees of freedom*

In this example, a spring-mass-damper system with three degrees of freedom (DOFs) modified from [75] is used for testing the capability of the proposed X-SVR based structural dynamic reliability analysis. The detailed configuration of the adopted vibration system is shown in Fig. 9. The dynamic equations of this spring-mass-damper system can be expressed as Eq. (57):

$$\begin{cases} m_1\ddot{u}_1(t)+(c_1+c_2)\dot{u}_1(t)-c_2\dot{u}_2(t)+(k_1+k_2)u_1(t)-k_2u_2(t)=F(t) \\ m_2\ddot{u}_2(t)-c_2\dot{u}_1(t)+(c_2+c_3)\dot{u}_2(t)-c_3\dot{u}_3(t)-k_2u_1(t)+(k_2+k_3)u_2(t)-k_3u_3(t)=0 \\ m_3\ddot{u}_3(t)-c_3\dot{u}_2(t)+(c_3+c_4)\dot{u}_3(t)+(k_3+k_4)u_3(t)-k_3u_2(t)=0 \end{cases} \quad (57)$$

where $m_1$, $m_2$ and $m_3$ denote the masses of the system, respectively; $k_1$, $k_2$, $k_3$ and $k_4$ denote the stiffness of the springs in the system, respectively; $c_1, c_2, c_3$ and $c_4$ denote the dampers of the system; $u_1(t), u_2(t)$ and $u_3(t)$ denote the displacements of the three lumped masses in this system. A time-dependent harmonic excitation $F(t)=F_0e^{-0.5t}\cos(4t)$ is applied to the mass $m_1$. The initial conditions of this system are set as $[u_1(0),u_2(0),u_3(0),\dot{u}_1(0),\dot{u}_2(0),\dot{u}_3(0)]=[0,0,0,0,0,0]$.

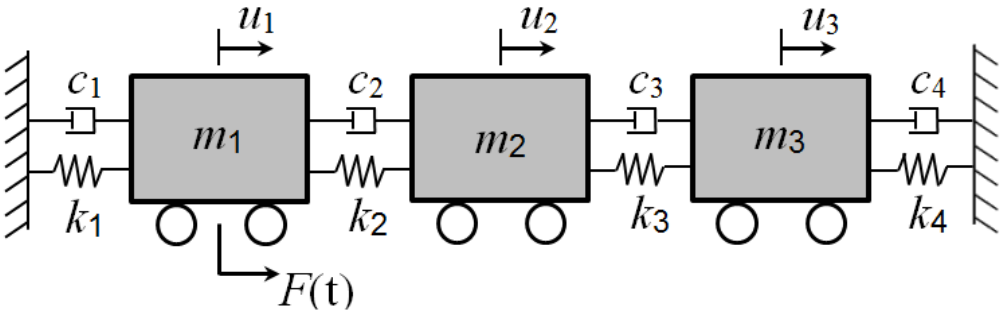



Fig 9. The spring-mass-damper system with three degrees of freedom (DOFs)

673 Table 4 The statistical information for uncertain parameters of Example 3

| Parameter | | $F_0$ (N) | $m_1, m_2, m_3$ (kg) | $c_1, c_2, c_3, c_4$ (N/(m/s)) | $k_1, k_2, k_3, k_4$ (N/m) |
|---|---|---|---|---|---|
| Mean ( $\mu$ ) | | 5 | 3 | 0.475 | 95 |
| Distribution ( $\sigma$ ) | | Normal | Lognormal | Lognormal | Lognormal |
| COV | Case 1 | 2.5% | | | |
| | Case 2 | 5% | | | |
| | Case 3 | 10% | | | |

674

675 The amplitude $F_0$ of the harmonic excitation is considered as a random variable with normal
676 distribution while the masses, stiffness and dampers of the system are considered as random
677 variables with lognormal distribution. Moreover, three different coefficients of variation (COVs)
678 ranging from 2.5% to 10% are studied in this example. The details of the statistical information of
679 the considered 12 random variables are listed in Table 4. For the investigation, the displacement of
680 $m_1$ is assumed to be critical for the safety of the system. According to the first-passage failure theory,
681 the limit state function of the spring-mass-damper system is defined by

682 $$g(\mathbf{x}) = \overline{u_{m_1}} - \max_t(|u_1(t)|) \tag{58}$$

683 where $\mathbf{x} = [k_1, k_2, k_3, k_4, c_1, c_2, c_3, c_4, m_1, m_2, m_3, F_0]$ and $\overline{u_{m_1}}$ denotes the allowable displacement at $m_1$
684 and is set to be 8.5 cm. The COV is defined as Eq. (59):

685 $$COV = \frac{\sigma_{input}}{\mu_{input}} \tag{59}$$

686 where $\mu_{input}$ and $\sigma_{input}$ denote the mean and standard deviation of the input random variable,
687 respectively.

688 The "exact" probability of failure ( $\tilde{p}_f$ ) for the considered three cases is obtained by using direct
689 Monte Carlo simulation with $N_{MCS} = 10^6$ samplings. For demonstrating the capability of the
690 proposed surrogate model, the X-SVR models are respectively constructed with various training
691 samples ( $N_{train} = 25, 50, 100$ ) and the predicted probability of failure ( $\hat{p}_f$ ) is then calculated
692 accordingly. Meanwhile, the conventional ε-SVR model with Gaussian kernel is trained and tested
693 with the same datasets. The accuracy of the proposed surrogate model is measured by calculating the
694 relative difference $\varepsilon_{p_f}$ between $\hat{p}_f$ and $\tilde{p}_f$ as the Eq. (60).

$$\varepsilon_{p_f} = \frac{\left|\hat{p}_f - \tilde{p}_f\right|}{\hat{p}_f} \times 100 \tag{60}$$

Table 4 The probability of failure – Case 1, Example 3

| Method | Case 1 | | | | | |
|---|---|---|---|---|---|---|
| | $\hat{p}_f$ ($N_{train} = 25$) | $\varepsilon_{p_f}$ | $\hat{p}_f$ ($N_{train} = 50$) | $\varepsilon_{p_f}$ | $\hat{p}_f$ ($N_{train} = 100$) | $\varepsilon_{p_f}$ |
| X-SVR | 0.1006 | 12.25% | 0.0808 | 9.86% | 0.0861 | 3.90% |
| ε-SVR | 0.1224 | 36.51% | 0.0570 | 36.46% | 0.1138 | 26.97% |
| MCS | $\tilde{p}_f = 0.0896$ | | | | | |

Table 5 The probability of failure – Case 2, Example 3

| Method | Case 2 | | | | | |
|---|---|---|---|---|---|---|
| | $\hat{p}_f$ ($N_{train} = 25$) | $\varepsilon_{p_f}$ | $\hat{p}_f$ ($N_{train} = 50$) | $\varepsilon_{p_f}$ | $\hat{p}_f$ ($N_{train} = 100$) | $\varepsilon_{p_f}$ |
| X-SVR | 0.3160 | 12.39% | 0.2924 | 3.99% | 0.2903 | 3.25% |
| ε-SVR | 0.2123 | 24.49% | 0.3017 | 7.29% | 0.2961 | 5.31% |
| MCS | $\tilde{p}_f = 0.2812$ | | | | | |

The simulated probability of failure by using the proposed method and ε-SVR model for Cases 1-3 are summarised in Table 4-6, respectively. By reviewing the simulation results, the proposed method surpasses the classical support vector regression model by offering less relative difference in predicting the probability of failure of the investigated spring-mass-damping system under various uncertainty levels. Additionally, it is indicated by the results that the probability of failure estimated by the proposed X-SVR approaches to $\hat{p}_f$ with increasing number of samples for training, while such trend is not obvious based on the simulation results obtained from ε-SVR.

Table 6 The probability of failure – Case 3, Example 3

| Method | Case 3 | | | | | |
|---|---|---|---|---|---|---|
| | $\hat{p}_f$ ($N_{train} = 25$) | $\varepsilon_{p_f}$ | $\hat{p}_f$ ($N_{train} = 50$) | $\varepsilon_{p_f}$ | $\hat{p}_f$ ($N_{train} = 100$) | $\varepsilon_{p_f}$ |

| | | | | | | |
|---|---|---|---|---|---|---|
| X-SVR | 0.3569 | 7.48% | 0.3641 | 5.58% | 0.3978 | 3.15% |
| ε-SVR | 0.2458 | 36.28% | 0.2728 | 29.28% | 0.1827 | 52.62% |
| MCS | $\tilde{p}_f = 0.3857$ | | | | | |

*5.4 Acoustic wave radiation from a 3D open structure*

The effectiveness and capability of the proposed X-SVR based reliability analysis approach is further applied to acoustic problem in this section. The 3D acoustic analysis of an open structure submerged in an infinite acoustic space is investigated. As shown in Fig. 10, the open structure is assumed to be a rigid hollow sphere with one quarter cut-off, where $r_0$ and $r_1$ ($r_1 = 1.2r_0$) denote the inner and outer radii of the open hollow sphere, respectively. The inner surface of the open sphere is uniformly subjected to a transient acoustic flux $f_s(t)$, which is defined in a dimensionless manner in Fig. 11. In Fig. 11(a), the applied transient flux is normalized by its peak magnitude $F_{st}$ and plotted against the dimensionless time $t_d = t\dfrac{\hat{c}}{r_0}$, where $\hat{c}$ denotes the nominal value of sound speed $c$. The Fourier transform of the flux $f_s(t)$ is also shown in Fig. 11(b) with the dimensionless amplitude $F_s(\omega)/F_{st}$ and the dimensionless frequency $\omega_d = \omega r_0/\hat{c}$. The corresponding highest frequency of interest is estimated to be $\omega_{d,\max} \approx 6$.

In this example, the time-dependent acoustic pressure $p_O(t)$ at point $O(0,0,0)$ marked in Fig. 10 is considered for reliability analysis. The commercial software ANSYS is employed for the acoustic analysis. Due to the infinitely large acoustic field, the acoustic field is firstly truncated to a bounded and an unbounded acoustic domain by the translucent spherical shell in Fig. 10 with the radius of $r_2$. In the ANSYS model, the bounded acoustic domain is modelled by FLUID30 elements (4-node tetrahedral elements). The unbounded acoustic domain is represented by the FLUID130 elements (3-node triangular elements) attached to the outer surface (spherical surface) of the bounded domain. According to Fig. 11(b), the minimum wavelength can be calculated as $\lambda_{\min} = \dfrac{2\pi}{\omega_{d,\max}} r_0 \approx 1.05 r_0$. In order to guarantee the accuracy, the element edge length is set to be $0.1r_0$ in ANSYS. This can provide approximately 11 nodes per minimum wavelength. From the ANSYS manual [76], the FLUID130 element is recommended to be placed at $0.2\dfrac{2\pi}{\omega_0} r_0 \approx 1.88 r_0$ away from the source of excitation. Therefore, the radius if the truncated spherical boundary $r_2$ is set to be $3r_1$ in

this model for accuracy.The total number of elements and nodes in this acoustic model are 1566852 and 271172, respectively. The mesh of this acoustic model is also illustrated in Fig. 12. The global equations of motion for this acoustic model can be formulated as:

$$\mathbf{M}_G^a \ddot{\mathbf{z}}_G(t) + \mathbf{C}_G^a \dot{\mathbf{z}}_G(t) + \mathbf{K}_G^a \mathbf{z}_G(t) = \mathbf{r}_G(t) \tag{61}$$

where $\mathbf{M}_G^a$, $\mathbf{C}_G^a$ and $\mathbf{K}_G^a$ denote the global mass, damping and stiffness matrices of acoustics, respectively; $\mathbf{z}_G(t)$ is the vector containing all nodal pressures $\mathbf{p}(t)$; and $\mathbf{r}_G(t)$ denotes the global flux vector which relates to $f_s(t)$. It should be noticed that the existence of damping matrix in Eq.(61) is due to the unbounded acoustic domain. The initial condition of the system is $[\mathbf{z}_G(0), \dot{\mathbf{z}}_G(0)] = [\mathbf{0}, \mathbf{0}]$.

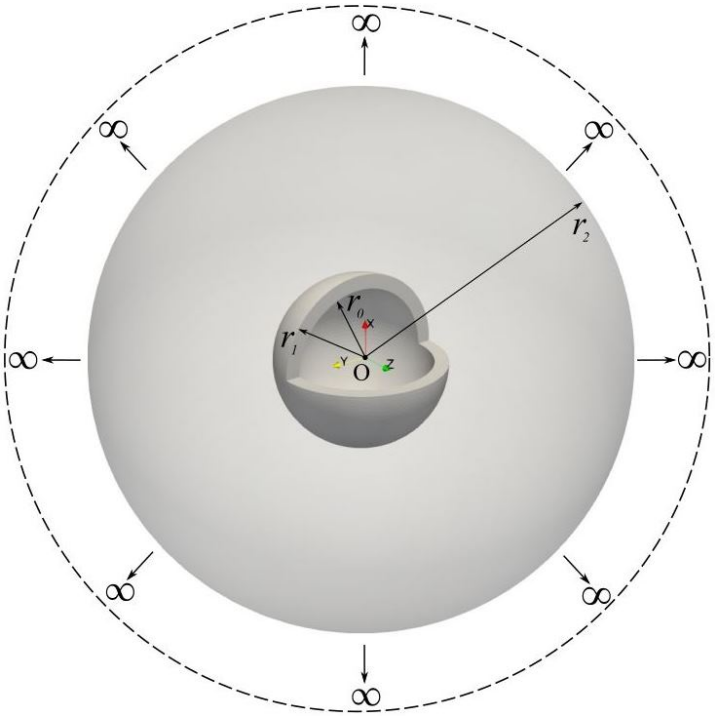
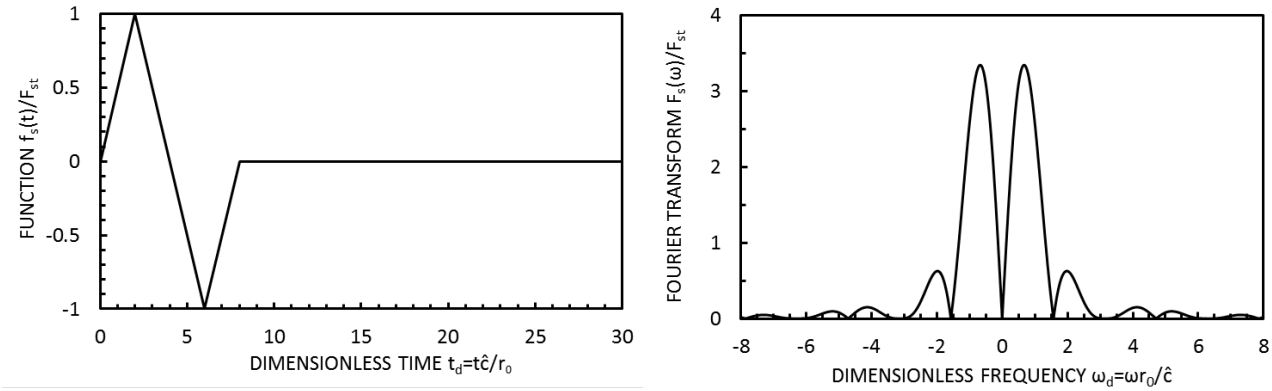



Fig. 10. Problem setup of a 3D open-sphere structure submerged in infinite acoustic domain.




(a) (b)

Fig. 11. Time variation of transient flux $f_s(t)$ applied on the inner spherical boundary: (a) time history and (b) Fourier transform.

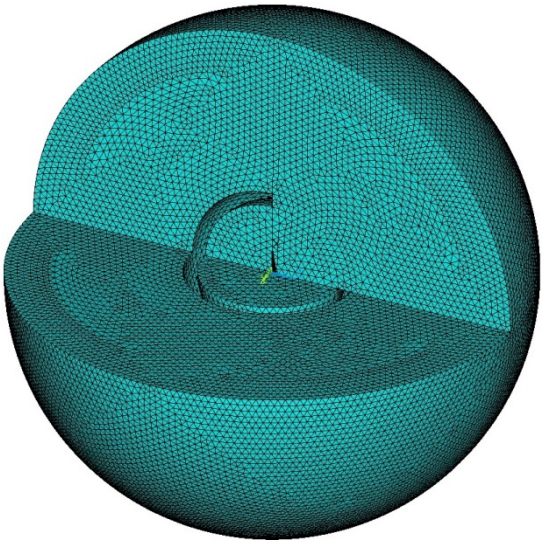


Fig. 12. The mesh used in ANSYS for modelling the acoustic field of the 3D open-structure model.

The sound speed $c$ and the peak flux magnitude $F_{st}$ are considered as random variables following Gaussian distribution with $COV = 10\%$ . For the purpose of demonstrating the effectiveness of the proposed method, the input parameters are considered as dimensionless [77] while $\hat{c}$ and the nominal value of $F_{st}$ are both defined as unitary such that $\mu_c = 1$ and $\mu_{F_{st}} = 1$. The uncertainties in $c$ and $F_{st}$ results in the fluctuation in the pressure response, which can be demonstrated in Fig. 13. The limit state of the acoustic system is defined as the dimensionless pressure $\frac{p_O(t)}{r_0 F_{st}}$ at point $O(0,0,0)$ shall not exceed the ultimate capacity which is assumed as $\overline{\hat{p}_O} = 8.5$ in this example. Thus, the limit state function for this example can be expressed as in Eq. (62).

$$g(c, F_{st}) = \overline{\hat{p}_O} - \max\left(\left|\frac{p_O(t)}{r_0 F_{st}}\right|\right) \tag{62}$$

Table 7. The probability of that the acoustic pressure at Point O exceeds $\overline{\hat{p}_O}$

| Method | Probability of failure | | | | | |
|---|---|---|---|---|---|---|
| | $\hat{p}_f$ $(N_{train} = 20)$ | $\varepsilon_{p_f}$ | $t_{com}$ (sec) | $\hat{p}_f$ $(N_{train} = 40)$ | $\varepsilon_{p_f}$ | $t_{com}$ (sec) |
| X-SVR | 0.1030 | 2.83% | 5.34 | 0.1040 | 1.89% | 4.53 |
| ε-SVR | 0.1140 | 7.55% | 6.87 | 0.1000 | 5.66% | 7.05 |
| MCS | $\tilde{p}_f = 0.1060$ | | | | | |

The X-SVR is employed for approximating the relationship between the input variables and extremum of the dimensionless pressure at point $O$. The probability of failure will be computed by using the constructed X-SVR surrogate model, which requires significantly less computational efforts than using the original ANSYS model. Similar to Example 3, the results obtained by the X-SVR model are comparing with the results obtained using ε-SVR with Gaussian kernel and the conventional MCS. Due to the excessive complexity of the model, the MCS is conducted with 1000 samples ( $N_{train}=1000$ ), which takes approximately $1.34\times10^{6}$ seconds. The predicted probability of failure ( $\hat{p}_f$ ) estimated by both X-SVR and ε-SVR and the 'exact' probability of failure ( $\tilde{p}_f$ ) obtained by the MCS ( $N_{train}=1000$ ) are summarized in Table 7. Moreover, the training time ( $t_{com}$ ) for constructing both the ε-SVR and X-SVR surrogate models is shown in Table 7, which indicates that the two methods require similar computational time. For the given numbers of training samples, the $\hat{p}_f$ obtained by X-SVR has less relative difference with $\tilde{p}_f$ in comparison with the ε-SVR. Furthermore, the probability density functions (PDFs) and cumulative distribution functions (CDFs) predicted by both X-SVR and ε-SVR are shown in Fig. 14 with the PDFs and CDFs obtained by the MCS. The kernel density estimation (KDE) which is a non-parametric approach to represent the PDFs and CDFs of random variables based on the available samples. Similar as the predicted probability of failure, the PDFs and CDFs obtained by the proposed X-SVR have relatively less variation to the ones obtained by the MCS, which is more visible in the PDF plots in Figs. 14(a) and 14(c). This study indicates that the proposed X-SVR model shows high efficiency and curacy for the reliability analysis of 3D acoustic application with unbounded domain.

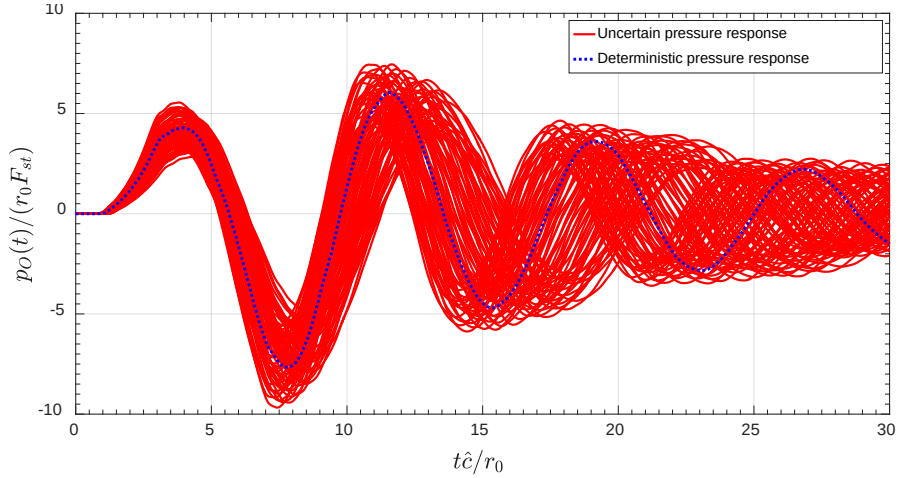
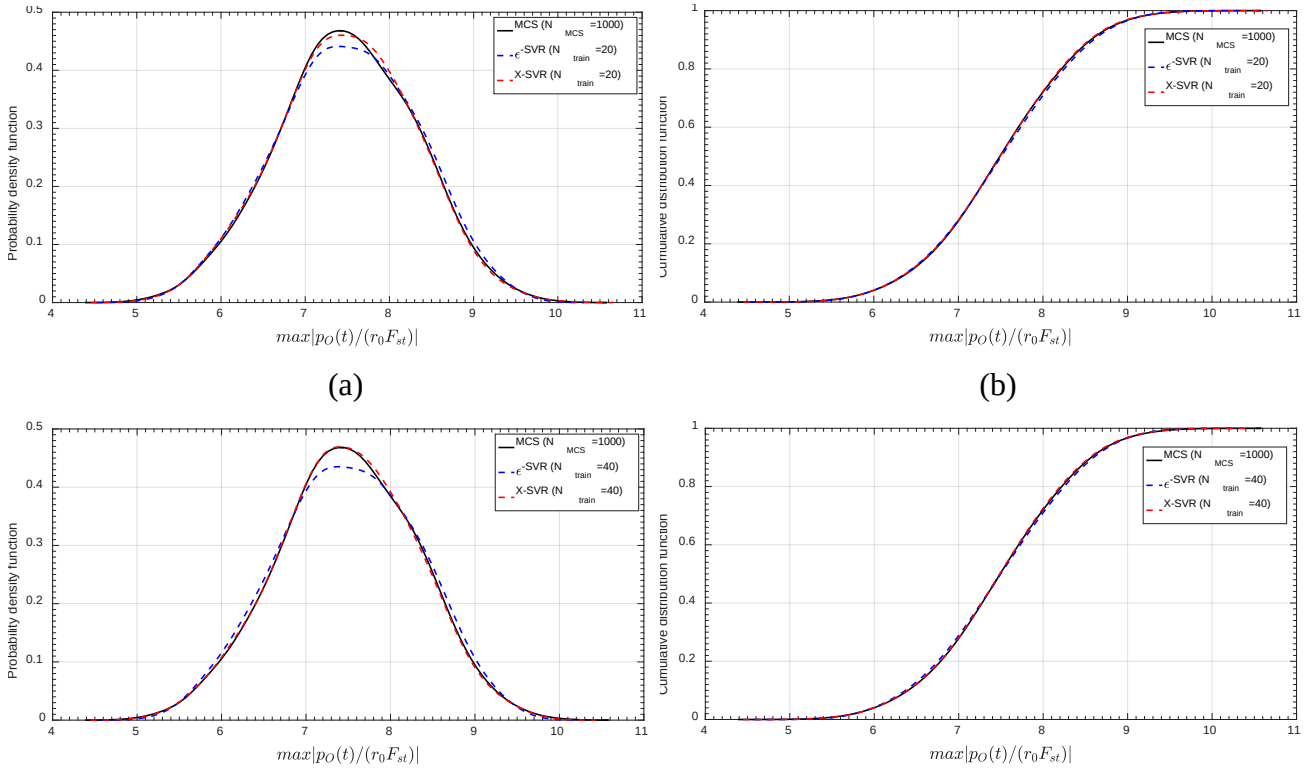



Fig.13. Uncertain acoustic pressure response (dimensionless) at point $O$ of the 3D open-structure model.




(a) (b)

(c) (d)

Fig. 14. The estimated PDFs and CDFs of $\max(|p_O(t)/r_0F_{st}|)$ obtained by X-SVR and ε-SVR with different $N_{train}$ and comparing with MCS

## 6. Conclusion

In this paper, a metamodel-based MCS strategy is proposed for the dynamic analysis with random input variables by evaluating the first-passage failure probability of systems. Within the proposed framework, the extended support vector regression (X-SVR) is introduced based on the theory of doubly regularized support vector machine. Since the proposed model can be formulated as convex quadratic programming problem, the global optimal solution for the given training dataset is promised. The suitable X-SVR parameters can be automatically selected by adopting the Bayesian optimization method. To enhance the capability of the introduced X-SVR approach, a new orthogonal polynomial kernel function satisfying the Mercer's condition is proposed by vectorizing the Gegenbauer polynomial. By implementing the proposed approach, an explicit function is constructed by training the X-SVR model to approximate the relationship between the input uncertain parameters and the extremum dynamic response of the system within a given time interval. Subsequently, the limit state function of the system can be efficiently evaluated such that the computational efficiency for obtaining the probability of failure using the MCS can be increased. The feasibility, efficiency and capability of the proposed method are systematically investigated by utilizing two benchmark examples and two engineering problems. by Comparing the results obtained

by proposed X-SVR model, the ε-SVR with Gaussian kernel and conventional MCS, the superior performance of the proposed method is evidently demonstrated.

A further extension of the proposed X-SVR based dynamic reliability analysis approach is to combine it with the advanced sample methods. Therefore, an adaptive X-SVR model can be constructed. Additionally, increasing the efficiency in solving the optimization problem will also be included in the future work.


## Acknowledgement

The work presented in this paper has been supported by Australian Research Council project DP160103919 and IH150100006.

The authors would like to sincerely express our gratitude to the anonymous reviewers for their valuable comments and constructive suggestions.


## Appendix A. Proof of the Proposition 1

The proof of the *Proposition 1* is demonstrated in detail as following.

*Proof*:

For quadratic programming expressed in Eq. (32), the prove of convexity is equivalent to proving that $\mathbf{Q} \succeq 0$. Moreover, considering that $\hat{\mathbf{C}}$ is a positive and diagonal matrix by definition, then $\hat{\mathbf{C}} \succ 0$ and also $\hat{\mathbf{C}}^{-1} \succ 0$. Let $\mathbf{v} \in \Re^{2m+2n}$ be a non-zero column vector, then:

$$\begin{aligned} \mathbf{v}^T\mathbf{Q}\mathbf{v} &= \mathbf{v}^T((\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})\hat{\mathbf{C}}^{-1}(\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})^T+\hat{\mathbf{G}}\hat{\mathbf{e}}\hat{\mathbf{e}}^T\hat{\mathbf{G}})\mathbf{v} \\ &= ((\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})^T\mathbf{v})\hat{\mathbf{C}}^{-1}((\hat{\mathbf{A}}+\mathbf{I}_{(2m+2n)\times(2m+2n)})^T\mathbf{v})+(\hat{\mathbf{e}}^T\hat{\mathbf{G}}\mathbf{v})^2 \\ &\geq 0 \end{aligned} \tag{A1}$$

Therefore, the dual problem defined in Eq. (32) is convex quadratic programming problem.

This concludes the proof. □

## Appendix B. Proof of the Proposition 2

The proof of the *Proposition 2* is demonstrated in detail as following.

*Proof*:

Firstly, the proposed GGK can be alternatively expressed as the product of two kernel functions $K_1(\mathbf{x}_i,\mathbf{x}_j)=\exp(-\sigma\left\|\mathbf{x}_i-\mathbf{x}_j\right\|_2^2)$ and $K_2(\mathbf{x}_i,\mathbf{x}_j)=\sum_{l=0}^{d}P_l^{\alpha_k}(\mathbf{x}_i)^T P_l^{\alpha_k}(\mathbf{x}_j)$ such that:

$$K_{GGK}(\mathbf{x}_i,\mathbf{x}_j)=K_1(\mathbf{x}_i,\mathbf{x}_j)K_2(\mathbf{x}_i,\mathbf{x}_j) \tag{B1}$$

According to [41, 64], the multiplication of two valid Mercer kernels is also a valid kernel function. Since that $K_1(\mathbf{x}_i,\mathbf{x}_j)$ is the Gaussian kernel ( $\sigma>0$ ) which satisfied the Mercer Theorem,

$K_{GGK}(\mathbf{x}_i, \mathbf{x}_j)$ can be proved as a valid kernel by verifying that $K_2(\mathbf{x}_i, \mathbf{x}_j)$ satisfies the Mercer Theorem.

Given an arbitrary squared integrable function $g(\mathbf{x})$ defined as $g: \Re^n \to \Re$ and assuming each element in $\mathbf{x}_i$ and $\mathbf{x}_j$ is independent with each other, then

$$\begin{aligned}
&\iint K_2(\mathbf{x}_i, \mathbf{x}_j) g(\mathbf{x}_i) g(\mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j \\
&= \iint \sum_{l=1}^{d} P_l^{\alpha_k}(\mathbf{x}_i)^T P_l^{\alpha_k}(\mathbf{x}_j) g(\mathbf{x}_i) g(\mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j \\
&= \sum_{l=0}^{d} \iint P_l^{\alpha_k}(\mathbf{x}_i)^T P_l^{\alpha_k}(\mathbf{x}_j) g(\mathbf{x}_i) g(\mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j \\
&= \sum_{l=0}^{d} \left[ \int P_l^{\alpha_k}(\mathbf{x}_i)^T g(\mathbf{x}_i) d\mathbf{x}_i \int P_l^{\alpha_k}(\mathbf{x}_j) g(\mathbf{x}_j) d\mathbf{x}_j \right] \\
&= \sum_{l=0}^{d} \left[ \left( \int P_l^{\alpha_k}(\mathbf{x}_i)^T g(\mathbf{x}_i) d\mathbf{x}_i \right) \left( \int P_l^{\alpha_k}(\mathbf{x}_i) g(\mathbf{x}_i) d\mathbf{x}_i \right) \right] \geq 0
\end{aligned} \tag{B2}$$

Thus, $K_2(\mathbf{x}_i, \mathbf{x}_j)$ is a valid Mercer kernel. Therefore, the proposed GGK $K_{GGK}(\mathbf{x}_i, \mathbf{x}_j)$ is an admissible Mercer kernel function.

This concludes the proof. □

## References


[1] Beer, M., Ferson, S. and Kreinovich, V., 2013. Imprecise probabilities in engineering analyses. Mechanical systems and signal processing, 37(1-2), 4-29.

[2] Chen, N., Yu, D., Xia, B. and Beer, M., 2016. Uncertainty analysis of a structural–acoustic problem using imprecise probabilities based on p-box representations. Mechanical Systems and Signal Processing, 80, 45-57.

[3] Wu, J., Zhang, Y., Chen, L. and Luo, Z., 2013. A Chebyshev interval method for nonlinear dynamic systems under uncertainty. Applied Mathematical Modelling, 37(6), 4578-4591.

[4] Muscolino, G., Sofi, A. and Giunta, F., 2018. Dynamics of structures with uncertain-but-bounded parameters via pseudo-static sensitivity analysis. Mechanical Systems and Signal Processing, 111, pp.1-22.

[5] Zhang, H., 2018. Durability reliability analysis for corroding concrete structures under uncertainty. Mechanical Systems and Signal Processing, 101, 26-37.

[6] Sofi, A. and Romeo, E., 2016. A novel interval finite element method based on the improved interval analysis. Computer Methods in Applied Mechanics and Engineering, 311, pp.671-697.

[7] Jiang, C., Ni, B.Y., Liu, N.Y., Han, X. and Liu, J., 2016. Interval process model and non-random vibration analysis. Journal of Sound and Vibration, 373, pp.104-131.

[8] Li, K., Gao, W., Wu, D., Song, C. and Chen, T., 2018. Spectral stochastic isogeometric analysis of linear elasticity. Computer Methods in Applied Mechanics and Engineering, 332, pp.157-190.

[9] Sofi, A., 2015. Structural response variability under spatially dependent uncertainty: stochastic versus interval model. Probabilistic Engineering Mechanics, 42, pp.78-86.

[10] Chen, J.B. and Li, J., 2007. The extreme value distribution and dynamic reliability analysis of nonlinear structures with uncertain parameters. Structural Safety, 29(2), 77-93.

[11] Jiang, C., Huang, X.P., Han, X. and Zhang, D.Q., 2014. A time-variant reliability analysis method based on stochastic process discretization. Journal of Mechanical Design, 136(9), p.091009.

[12] Jiang, C., Ni, B.Y., Han, X. and Tao, Y.R., 2014. Non-probabilistic convex model process: a new method of time-variant uncertainty analysis and its application to structural dynamic reliability problems. Computer Methods in Applied Mechanics and Engineering, 268, 656-676.

[13] Shi, Y., Lu, Z., Zhang, K. and Wei, Y., 2017. Reliability analysis for structures with multiple temporal and spatial parameters based on the effective first-crossing point. Journal of Mechanical Design, 139(12), 121403-121403.

[14] Du, X., 2014. Time-dependent mechanism reliability analysis with envelope functions and first-order approximation. Journal of Mechanical Design, 136(8), p.081010.

878 [15] Spanos, P.D, Kougioumtzoglou, I.A., 2014. Survival probability determination of nonlinear
879 oscillators subject to evolutionary stochastic excitation. Journal of Applied Mechanics, vol. 81, no. 5,
880 051016
881 [16] Kougioumtzoglou, I.A., Spanos, P.D., 2013. Nonlinear MDOF system stochastic response
882 determination via a dimension reduction approach. Computers and Structures, vol. 126, p. 135-148,
883 [17] Kougioumtzoglou, I.A., Spanos, P.D., 2013. Response and first-passage statistics of nonlinear
884 oscillators via a numerical path integral approach. ASCE Journal of Engineering Mechanics, vol. 139,
885 no. 9, p. 1207-1217
886 [18] Wang, Z. and Chen, W., 2016. Time-variant reliability assessment through equivalent stochastic
887 process transformation. Reliability Engineering & System Safety, 152, 166-175.
888 [19] Goller, B., Pradlwarter, H.J. and Schuëller, G.I., 2013. Reliability assessment in structural
889 dynamics. Journal of Sound and Vibration, 332(10), 2488-2499.
890 [20] Mitseas, I. P.; Kougioumtzoglou, I. A.; Spanos, P. D.; Beer, M., 2016. Nonlinear MDOF
891 structural system survival probability determination subject to evolutionary stochastic excitation,
892 Journal of Mechanical Engineering, 62(7–8), 440–451
893 [21] Li, J., Chen, J., Sun, W. and Peng, Y., 2012. Advances of the probability density evolution
894 method for nonlinear stochastic systems. Probabilistic Engineering Mechanics, 28, 132-142.
895 [22] Xu, J. and Kong, F., 2018. An adaptive cubature formula for efficient reliability assessment of
896 nonlinear structural dynamic systems. Mechanical Systems and Signal Processing, 104, 449-464.
897 [23] Andrieu-Renaud, C., Sudret, B. and Lemaire, M., 2004. The PHI2 method: a way to compute
898 time-variant reliability. Reliability Engineering & System Safety, 84(1), 75-86.
899 [24] Schueller, G.I., Pradlwarter, H.J., Koutsourelakis, P.S., 2004. A critical appraisal of reliability
900 estimation procedures for high dimensions. Probabilistic Engineering Mechanics, vol. 19, no. 4, p.
901 463-474.
902 [25] Wang, Z. and Wang, P., 2012. A nested extreme response surface approach for time-dependent
903 reliability-based design optimization. Journal of Mechanical Design, 134(12), p.121007.
904 [26] Lu, C., Feng, Y.W., Liem, R.P. and Fei, C.W., 2018. Improved Kriging with extremum response
905 surface method for structural dynamic reliability and sensitivity analyses. Aerospace Science and
906 Technology, 76, 164-175.
907 [27] Wang, Z. and Wang, P., 2015. A double-loop adaptive sampling approach for sensitivity-free
908 dynamic reliability analysis. Reliability Engineering & System Safety, 142, 346-356.
909 [28] Wu, J., Luo, Z., Zhang, Y., Zhang, N. and Chen, L., 2013. Interval uncertain method for
910 multibody mechanical systems using Chebyshev inclusion functions. International Journal for
911 Numerical Methods in Engineering, 95(7), pp.608-630.

[29] Wu, J., Luo, Z., Zhang, N. and Zhang, Y., 2015. A new interval uncertain optimization method for structures using Chebyshev surrogate models. Computers & Structures, 146, pp.185-196.
[30] Fei, C.W. and Bai, G.C., 2014. Distributed collaborative probabilistic design for turbine blade-tip radial running clearance using support vector machine of regression. Mechanical Systems and Signal Processing, 49(1-2), 196-208.
[31] Gao, W. and Kessissoglou, N.J., 2007. Dynamic response analysis of stochastic truss structures under non-stationary random excitation using the random factor method. Computer Methods in Applied Mechanics and Engineering, 196(25-28), 2765-2773.
[32] Xia, B., Yu, D. and Liu, J., 2013. Hybrid uncertain analysis for structural–acoustic problem with random and interval parameters. Journal of Sound and Vibration, 332(11), pp.2701-2720.
[33] Ma, J., Gao, W., Wriggers, P., Chen, J. and Sahraee, S., 2011. Structural dynamic optimal design based on dynamic reliability. Engineering Structures, 33(2), pp.468-476.
[34] Jiang, C., Bi, R.G., Lu, G.Y. and Han, X., 2013. Structural reliability analysis using non-probabilistic convex model. Computer Methods in Applied Mechanics and Engineering, 254, pp.83-98.
[35] Wu, D. and Gao, W., 2017. Hybrid uncertain static analysis with random and interval fields. Computer Methods in Applied Mechanics and Engineering, 315, 222-246.
[36] Melchers, R.E. and Beck, A.T., 2018. Structural reliability analysis and prediction. John Wiley & Sons.
[37] Ma, J., Gao, W., Wriggers, P., Wu, T. and Sahraee, S., 2010. The analyses of dynamic response and reliability of fuzzy-random truss under stationary stochastic excitation. Computational Mechanics, 45(5), 443-455.
[38] Ma, J., Wriggers, P., Gao, W., Chen, J.J. and Sahraee, S., 2011. Reliability-based optimization of trusses with random parameters under dynamic loads. Computational Mechanics, 47(6), pp.627-640.
[39] Wu, B., Wu, D., Gao, W. and Song, C., 2018. Time-variant random interval natural frequency analysis of structures. Journal of Sound and Vibration, 414, 284-298.
[40] Wu, J., Luo, Z., Zhang, N., Zhang, Y. and Walker, P.D., 2017. Uncertain dynamic analysis for rigid-flexible mechanisms with random geometry and material properties. Mechanical Systems and Signal Processing, 85, 487-511.
[41] Vapnik, V., 2013. The nature of statistical learning theory. Springer science & business media.
[42] Smola, A.J. and Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and computing, 14(3), 199-222.
[43] Kung, S.Y., 2014. Kernel methods and machine learning. Cambridge University Press.

946 [44] Wang, L., Zhu, J. and Zou, H., 2006. The doubly regularized support vector machine. Statistica
947 Sinica, 16(2), p.589.
948 [45] Zou, H. and Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of
949 the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.
950 [46] Dunbar, M., Murray, J.M., Cysique, L.A., Brew, B.J. and Jeyakumar, V., 2010. Simultaneous
951 classification and feature selection via convex quadratic programming with application to HIV-
952 associated neurocognitive disorder assessment. European Journal of Operational Research, 206(2),
953 470-478.
954 [47] Fung, G.M. and Mangasarian, O.L., 2004. A feature selection Newton method for support vector
955 machine classification. Computational optimization and applications, 28(2), 185-202.
956 [48] Jeyakumar, V., Li, G. and Suthaharan, S., 2014. Support vector machine classifiers with
957 uncertain knowledge sets via robust optimization. Optimization, 63(7), 1099-1116.
958 [49] Portera, F. and Sperduti, A., 2005. Support vector regression with a generalized quadratic loss.
959 In Biological and Artificial Intelligence Environments (pp. 209-216). Springer, Dordrecht.
960 [50] Mangasarian, O.L. and Musicant, D.R., 2001. Lagrangian support vector machines. Journal of
961 Machine Learning Research, 1(Mar), pp.161-177.
962 [51] Scholkopf, B., Mika, S., Burges, C.J., Knirsch, P., Muller, K.R., Ratsch, G. and Smola, A.J.,
963 1999. Input space versus feature space in kernel-based methods. IEEE transactions on neural
964 networks, 10(5), pp.1000-1017.
965 [52] Gao, J., Kwan, P.W. and Shi, D., 2010. Sparse kernel learning with LASSO and Bayesian
966 inference algorithm. Neural Networks, 23(2), 257-264.
967 [53] Feng, Y., Lv, S.G., Hang, H. and Suykens, J.A., 2016. Kernelized elastic net regularization:
968 Generalization bounds, and sparse recovery. Neural computation, 28(3), pp.525-562.
969 [54] Dai, H., Zhang, B. and Wang, W., 2015. A multiwavelet support vector regression method for
970 efficient reliability assessment. Reliability Engineering & System Safety, 136, 132-139.
971 [55] Su, H., Li, X., Yang, B. and Wen, Z., 2018. Wavelet support vector machine-based prediction
972 model of dam deformation. Mechanical Systems and Signal Processing, 110, pp.412-427.
973 [56] Wu, Q., 2011. Hybrid wavelet ν-support vector machine and chaotic particle swarm optimization
974 for regression estimation. Expert Systems with Applications, 38(12), pp.14624-14632.
975 [57] Ozer, S., Chen, C.H. and Cirpan, H.A., 2011. A set of new Chebyshev kernel functions for
976 support vector machine pattern classification. Pattern Recognition, 44(7), pp.1435-1447.
977 [58] Cheng, K., Lu, Z., Wei, Y., Shi, Y. and Zhou, Y., 2017. Mixed kernel function support vector
978 regression for global sensitivity analysis. Mechanical Systems and Signal Processing, 96, pp.201-214.

[60] Tian, M. and Wang, W., 2017. Some sets of orthogonal polynomial kernel functions. Applied Soft Computing, 61, pp.742-756.
[61] Yin, S., Yu, D., Yin, H. and Xia, B., 2016. Interval and random analysis for structure–acoustic systems with large uncertain-but-bounded parameters. Computer Methods in Applied Mechanics and Engineering, 305, pp.910-935.
[62] San Kim, D., Kim, T. and Rim, S.H., 2012. Some identities involving Gegenbauer polynomials. Advances in Difference Equations, 2012(1), p.219.
[63] Padierna, L.C., Carpio, M., Rojas-Domínguez, A., Puga, H. and Fraire, H., 2018. A novel formulation of orthogonal polynomial kernel functions for SVM classifiers: The Gegenbauer family. Pattern Recognition, 84, pp.211-225.
[64] Gönen, M. and Alpaydın, E., 2011. Multiple kernel learning algorithms. Journal of machine learning research, 12(Jul), pp.2211-2268.
[65] Deng, N., Tian, Y. and Zhang, C., 2012. Support vector machines: optimization based theory, algorithms, and extensions. Chapman and Hall/CRC.
[66] Cheng, K. and Lu, Z., 2018. Adaptive sparse polynomial chaos expansions for global sensitivity analysis based on support vector regression. Computers & Structures, 194, pp.86-96.
[67] Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., Patwary, M., Prabhat, M. and Adams, R., 2015, June. Scalable bayesian optimization using deep neural networks. In International Conference on Machine Learning (pp. 2171-2180).
[68] Snoek, J., Larochelle, H. and Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. In Advances in neural information processing systems (pp. 2951-2959).
[69] MathWorks, 2017. Statistical and machine learning toolbox: user's guide (2017b), Math Works Inc.
[70] Pan, Q. and Dias, D., 2017. An efficient reliability method combining adaptive support vector machine and Monte Carlo simulation. Structural Safety, 67, pp.85-95.
[71] Moustapha, M., Bourinet, J.M., Guillaume, B. and Sudret, B., 2018. Comparative Study of Kriging and Support Vector Regression for Structural Engineering Applications. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 4(2), p.04018005.
[72] Sobol', I.Y.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Computational Mathematics and Mathematical Physics, 7(4), pp.784-802.
[73] Kersaudy, P., Sudret, B., Varsier, N., Picon, O. and Wiart, J., 2015. A new surrogate modeling technique combining Kriging and polynomial chaos expansions–Application to uncertainty analysis in computational dosimetry. Journal of Computational Physics, 286, pp.103-117.

[74] Chatterjee, T. and Chowdhury, R., 2017. An efficient sparse Bayesian learning framework for stochastic response analysis. Structural Safety, 68, pp.1-14.
[75] Xia, B., Qin, Y., Yu, D. and Jiang, C., 2016. Dynamic response analysis of structure under time-variant interval process model. Journal of Sound and Vibration, 381, pp.121-138.
[76] ANSYS, Inc., 2017, ANSYS Mechanical APDL Structural Analysis Guide, Release 17.0, ANSYS, Canonsburg, Pennsylvania, US
[77] Birk, C., Liu, L. and Song, C., 2016. Coupled acoustic response of two-dimensional bounded and unbounded domains using doubly-asymptotic open boundaries. Journal of Computational Physics, 310, pp.252-284.